

Pure science in an age of attrition

Despite stupendous advances in the past 50 years, Western governments' support of research for its own sake is on the wane. The new year is an appropriate time for researchers and policy-makers to consider the implications and make some resolutions.

LONG gone are the days when society perceived the scientist as someone pursuing simply and exclusively, to use Samuel Taylor Coleridge's phrase, the "gratification of knowing". The 50 years since the end of the Second World War, especially, have seen continual tension between scientists and politicians about the linkage between the exploration of natural (and sometimes unnatural) phenomena on the one hand and their countries' interests on the other. The multinational survey of the period in this issue (see pages 5–9) reflects some of those tensions. But a look forward suggests that the lessons of history are not sufficient to cope with problems that are to come, and that original and sometimes unpleasant thinking is required on behalf of pure science in particular.

Regrettably, there are moral overtones to the word "pure" — "basic" is more neutral. But those who attain fulfilment merely from asking careful questions of the world and sometimes having them interestingly answered — at best, in a totally unexpected fashion — know the essential purity of the experience. With that comes a motivation, if not a compulsion, quite distinct from the determination to invent, to create wealth or to enhance the quality of life. But if scientists then find that they have new knowledge or experience to exploit, and enjoy the prospect of moving out of research to enrich society (and perhaps themselves) and leave openings for other bright young things, all well and good.

The fact that most good scientists do not behave like that gives rise to acute pain and much distraction if the budget for pure science is contracting — which, in too many countries, it is. In Japan and Germany funds are officially expanding, but governments are finding it hard to deliver on their promises. But there are many countries where researchers are finding more and more frequently that high quality, non-strategic ideas that are not immediately applicable fail to attract funds. This trend will continue for some time to come, given the increasingly stressed state of most industrial economies, combined with a correct but imbalanced perception by governments that strategic and immediately applicable ideas are important for international competitiveness.

The central political issues for scientists engaged in research for its own sake are therefore twofold: first, to fight back by seeking to ensure that governments' perceptions of the importance of their work are well founded; and second, to minimize the damage inflicted by declining resources. Although these challenges are not new to some disciplines, the pressures to come necessitate more radical responses than hitherto.

The first challenge requires greater cohesion between scientists, economists and analysts of science and technology policy, not least by talking to each other more. Mutual awareness is at present too often fragmentary. Improved contacts are necessary in order to illuminate the benefits to a modern economy of vigorous activities in pure sciences — benefits in skills and creative flexibility of key portions of the population, in education, in attracting inward investment, past benefits that were unanticipated products of basic research, and benefits in culture (which also contribute positively to economies).

This is not to request a bending of economics and policy research to the scientists' policy agenda, but rather to make the most of them. And presenting a case for basic science in a sophisticated economic context can only strengthen its call on politicians' attention — especially if weaknesses in assumptions underlying an imbalance in support for applied science are also exposed. An important side issue here is to ensure that statistics of support for distinctively basic research are adequately maintained.

Fighting back also requires the development of political skills, so as to present the case more effectively than at present. Last year, for example, saw US biologists outshine physicists in protecting their budgets from political onslaught. Physicists were belatedly urged by professional bodies to write to their representatives in Congress. More generally, scientists need to question whether they can be sufficiently represented by learned societies in today's politics (to which the answer is no) and to do more to channel complementary representation through respectable lobby groups, funded from scientists' own pockets, and, even more than now, through the media.

The second challenge confronts basic researchers with a need for more collaborative funding and coordination. In a contracting financial universe, costly particle physics and space-based sciences especially face a need to sacrifice national aspirations to the overriding interests of internationalism. It is worrying that, with research and development towards the world's next-but-one major accelerator well under way, and despite the efforts of the OECD's Megascience Forum, decisions to build intermediate-size facilities (such as B-boson factories) are still being taken at a national level without adequate international planning. Similarly Japan, Europe and the United States urgently need to perceive their futures in space-based sciences as more fundamentally collaborative than has ever been the case in the past.

'Small science' will also require more internationalism in a way that challenges the interests of national funding agencies. In pure science, the people who need to be funded are simply the best, wherever they are. It is better, therefore, to ensure that, for example, several European countries support two excellent research centres in a hopelessly inapplicable patch of basic science than to allow the activity to fade out of Europe altogether. The need to institutionalize collaboration in European science funding well beyond the activities of the European Commission has never looked stronger.

Most challengingly of all, limiting the damage from declining funds also requires something of excellent scientists that is deeply impure: to sit on committees and, again and again, to be prepared to identify other excellent scientists who, for budgetary reasons only, must receive no more funds, occasionally in the process wiping out another subdiscipline within their country. No wonder some funding bodies have recently introduced anonymity of their peer-review panels. But however painful, such eliminations are best carried out by people who understand the damage they are causing — and who do not have to keep silent about it. □



EU-US cooperation talks may hinge on access to research programmes

Washington. The European Union (EU) and the United States have promised to negotiate a comprehensive science and technology cooperation agreement by 1997. But so far there is little agreement on what it should contain.

The promise was made as part of a 'New Transatlantic Agenda' endorsed at a summit meeting in Madrid on 3 December by Bill Clinton, the US president, Jacques Santer, president of the European Commission (EC), and Felipe Gonzalez, the Spanish prime minister, who headed the EU's Council of Ministers during Spain's presidency of the EU, from July to December 1995.

A statement issued after the summit said that the three participants "will negotiate a new comprehensive EC-US science and technology cooperation agreement by 1997 based on the principle of mutual interest, with a view to achieving a balance of benefits to us both".

Negotiations are expected to be tough as the United States and the EU have different expectations from an agreement. Officials at the commission would like European researchers to have better access to federally supported research programmes in the United States. Their US counterparts, on the other hand, want a narrower 'umbrella' agreement, primarily concerned with bringing patenting arrangements into line.

Concern about such 'institutional mismatching' between cooperating partners led to the rejection by the EU of a collaboration proposal offered by the United States 18 months ago. According to Jose Costa, science counsellor for the EC in Washington, this proposal was "an empty shell, which only allowed for consultation and for intellectual property rights".

Costa says that such an agreement would suit the United States, as it would leave the decentralized network of US research agencies to make their own deals with the Europeans. But the Europeans fear that this would allow the United States to gain access to Europe-wide research programmes while comparable US programmes funded by, for example, the Department of Energy, could remain closed at that agency's discretion.

"There is no systematic approach to this in the United States," says Costa. "You have everything from very open doors at the National Institutes of Health to others that are far more restrictive."

In March 1994, representatives of the EC and the European Parliament met in Brussels (see *Nature* 368, 385; 1994) to discuss a report — commissioned by the parliament's

Office of Scientific and Technological Operations Assessment — on technological cooperation with the United States and Japan.

Anne-Marie Goedmakers, a Dutch member of the European Parliament, said that it was unfair that European subsidiaries of US and Japanese companies could take part in EU research programmes, when subsidiaries of European companies were excluded from some US national programmes. The United States has already started to win access to some EC research programmes, such as the next phase of Esprit, the commission's information technology programme. US science agencies hope that any agreement will help to accelerate this process.

Jean Hudson, of the international programmes division of the National Science Foundation (NSF), says that she is enthusiastic about the prospect of any agreement that would improve transatlantic cooperation. "The United States is a large and wealthy country, and there are a lot of resources here, so US scientists are less likely than others to travel [abroad]. There have been huge strides in European countries' [scientific] capabilities, and we are not doing our job if we don't facilitate access there for our researchers."

Nevertheless, Hudson says, other factors are drawing scientists from the two regions apart. "On both sides, we feel that transatlantic ties are weakening," she says. "US scientists face a difficult job market, so they travel less. And European programmes encourage European scientists to stay in Europe."

Hudson adds that US scientists are excluded from some European-level

biotechnology and materials research programmes to which she would like them to have access. But the official US approach, led by the State Department, has been minimalist, merely proposing an agreement that would set rules for patenting, and leaving agencies such as the NSF to win access to European programmes later on.

"The access issue should be separated from this agreement," says Cathy Campbell at the White House's Office of Science and Technology Policy. "The agreement is to provide a framework for agencies to develop agreements with their counterparts in the European Union."

Officials on both sides say that the United States does not want to include the question of access in the talks, as its negotiators are ill-placed to deliver the cooperation of its diverse spectrum of research agencies. "We cannot change US law," says Campbell bluntly. "We have explained that to the European Union."

Officials at the European Commission must now draw up counter-proposals in order to win a mandate from the Council of Ministers to begin negotiations whose outcome is hard to predict. Despite the language of the communiqué, the officials say that the two sides are not committed to reach an agreement by 1997, but only to try to do so.

Some observers doubt the likelihood of any substantial agreement being reached. An official of one EU member country, who asked not to be named, said that the Madrid language was deliberately vague because officials had failed to agree anything more specific to put in the communiqué.

Colin Macilwain

1996 Fermi award honours his former student

London. An 83-year-old atomic physicist, whose theories helped develop the laser, and an 82-year-old chemist who discovered the radioisotope carbon-14, have been named by US president Bill Clinton as the winners of this year's Enrico Fermi award.

Italian-born Ugo Fano, professor emeritus at the University of Chicago and one of the last living students to have worked with Enrico Fermi, will receive the award with Canadian-born Martin Kamen, professor emeritus at the University of California at San Diego.

Kamen discovered the long-lived radioisotope in 1940 in collaboration with the late Sam Ruben. Carbon-14 today is used for applications including dating archaeo-

logical finds and tracking carbon dioxide in the environment. Fano's work on the interaction of radiation and matter has helped the development of nuclear medicine.

The 40-year-old award, which carries a \$100,000 honorarium and gold medal for each winner, is the US government's oldest science and technology award. It is granted in recognition for a lifetime's achievement in the field of nuclear energy.

The US president approved the winners from recommendations made by Hazel O'Leary, Secretary of Energy following an evaluation of candidates by a screening panel and an interagency awards committee. O'Leary will present the awards at a ceremony in Washington later this year. □

French risk assessment body faces uncertain future

Paris. The future of a commission set up by the French government to provide independent assessments of the risks associated with technology may be under threat, warned its chairman, Jean-Jacques Salomon, last week.

The so-called Collège de la Prévention des Risques Technologiques (CPRT), which was created in 1989 by the socialist prime minister, Michel Rocard, is attached to the office of the prime minister. Its 12 members are appointed by the president of the republic for six-year terms, and a third of the membership is renewed every two years.

The CPRT has two important powers. Although it produces 'opinions' at the request of the government, it can also do so on its own initiative. Second, the commission can make public its recommendations. It has published more than a dozen such opinions on industries such as nuclear power, chemicals and petroleum, and new technologies such as biotechnology.

But a rumour that the CPRT is to be abolished is now circulating. One reason is that the government is trying to save money — and the CPRT is perceived by some as an irritant. Moreover, although the independence of the commission is widely acknowledged, some in the conservative government are suspicious of a creation of the former socialist administration.

Circumstantial evidence of the govern-

ment's lack of interest comes from its failure to appoint a successor to Salomon, who resigned as chairman last year to concentrate on his work at the Conservatoire National des Arts et Métiers in Paris.

Salomon, who has continued as acting chairman of CPRT, admits that it is "heads or tails" that the CPRT will be abolished. The government's "temptation" to abolish the CPRT is being tempered, he says, by concern that such a move could further harm the reputation of a government that is already mistrusted by the public. "We will see in 1996 if there is provision made in the budget [for CPRT]."

The CPRT, which has an annual budget of just FF1.5 million (US\$300,000), is particularly vulnerable because it has no permanent staff. Another weakness that has compromised its effectiveness is that while in theory it has unlimited powers of investigation, in practice, there is no legal obligation on anybody to reply to its questions.

Nevertheless, Salomon argues that the abolition of CPRT would mean the loss of "an independent voice which calls for vigilance against technocrats". Indeed, it is widely acknowledged that France lacks the counterbalances to the power of government and industry that are common, for example, in the United States.

Declan Butler

Political uncertainty delays EMBL decision

Munich. A vote on plans to increase the budget of the Heidelberg-based European Molecular Biology Laboratory (EMBL) by six per cent over the next three years has been left open until mid-February because France was unable to make a decision at last month's EMBL council meeting.

All the other 14 member states have approved the budget increase which would be provided almost entirely by Germany, whose contributions to EMBL and the European Laboratory for Particle Physics (CERN) rise this year to take into account its increased economic weight following reunification. Member state contributions are broadly based on gross domestic product.

But France, caught between a recent reorganization of its research ministry (see *Nature* 378, 225; 1995) and general political uncertainty, is unable to confirm that it would be able to support the small increase

in its own contribution that would be demanded by the proposed budget.

EMBL is under considerable financial pressure as it tries to realize new scientific goals, already approved by council, set by its director Fotis Kafatos who took office nearly two years ago. These include the establishment of a new major programme in developmental biology at the main laboratory in Heidelberg, and expansion of two outstations, the European Bioinformatics Institute at Hinxton, near Cambridge, and the Grenoble outstation concerned with structural biology.

Plans also include an expanded visitors programme and the establishment of four EMBL research groups at a new centre for genetics at Monterotondo, near Rome, where the EU-sponsored mouse gene repository is to be sited.

Even if the budget is approved next month, Kafatos will still be faced with the prospect of achieving these ambitions on a level of funding that will barely keep up with inflation. But he remains optimistic: "everyone in Europe is facing financial hardship", he says, "and we will do our part".

Alison Abbott



Kafatos: still hoping

Natural History Museum joins Kew for lottery cash

London. The Royal Botanic Gardens at Kew and the Natural History Museum in London have secured substantial funds from the proceeds of Britain's National Lottery for two major projects.

The museum has been awarded £6 million (US\$9 million) from the National Lottery Heritage Fund towards its £12-million scheme to transform the former Geological Museum into a three-floor Earth sciences museum (see *Nature* 378, 2; 1995).

Kew Gardens, meanwhile, has won £21 million (US\$39 million) from the Millennium fund — set aside from lottery income for projects marking the beginning of the next century — for the Millennium Seed Bank, which will be built near Wakehurst Place in West Sussex, the site of Kew's existing facility. The new institution is expected to conserve at least 10 per cent of the world's flora by the year 2010.

Phase one of the Earth sciences gallery at the Natural History Museum is due to open in July this year. Three further exhibitions, part of phase two and three, are expected to open in the summer of 1998. □

EC encourages science for profit

Munich. The European Commission has published a green paper (discussion document) on innovation intended to encourage greater European competitiveness in converting scientific performance into profitability.

An initiative of the research commissioner, Edith Cresson, and the commissioners for industrial affairs and energy, the paper points to the low level of investment in research and development in Europe (2 per cent of gross national product) compared to Japan (2.7 per cent) and the United States (2.7 per cent), the slow and expensive procedures of patenting, and extensive bureaucracy as the main contributors to Europe's relative lack of innovation.

It calls for comments from scientists, industry and investors on 13 proposals for corrective action, which include the orientation of research at national and European levels towards applications, more programmes to encourage mobility of researchers, increased support for venture capital and tax benefits to encourage investment in research.

The commission is to consider producing a white paper (policy document) to be presented to the European Parliament and the Council of Ministers, based on the green paper and taking into account comments received during this year. **Alison Abbott**

After fifty years, a mixed verdict

"THE general public has not been unaware of, or ungrateful for, the contributions made by science to the task of defeating the enemy," wrote *Nature* the week after the cessation of hostilities with Germany in May 1945 (155, 553; 1945). "Indeed, there have been times when it has been necessary to check public enthusiasm and to remind people at large that science alone could not achieve [this result]."

Nevertheless, scientists were quick to exploit this outburst of enthusiasm — which, significantly, existed well before the public had become aware of the contribution of science to the development of the atomic bomb. Once the war had ended, plans were rapidly laid for schemes to ensure that both basic and applied science played as vital a part in post-war reconstruction as it had in war-time. These plans found their most articulate expression in *Science — The Endless Frontier*, an evocative and influential report presented by Vannevar Bush, director of the Office of Scientific Research and Development, to President Harry Truman.

But the United States was not alone. In France, substantial state support for basic research was channelled through the Centre Nationale de la Recherche Scientifique. In Germany, thanks partly to the support of British colleagues, nuclear physicists who

had opposed Hitler were able to oversee the foundations of the highly successful system of Max Planck Institutes.

Fifty years on, however, as this brief survey contrasting the worlds of 1945 and 1995 demonstrates, the impact of the powerful post-war legacy on science remains mixed. On the one hand, science has kept much of its social promise, providing the advances in fields such as materials research, information technology and biotechnology that form the backbone of modern economies.

In turn, governments have kept much of their promise to basic science. Until recently, academic research has grown steadily, while decisions on the distribution of these funds has remained largely in the hands of the scientific community.

But there are signs that the post-war paradigm is breaking up. The vital role of science in the economy remains accepted, but further growth in basic research is no longer guaranteed by governments, which are increasingly keen to see 'strategic priorities' imposed on the science base.

At the same time, in virtually all industrialized nations, many of the massive state-funded laboratories set up in the 1940s and 1950s to pursue social goals — in particular, the military and civilian applications of nuclear power — have become political

albatrosses, deprived of their original political and technological goals, but lacking the flexibility to respond speedily to the demands of the modern industrial world.

Nor is the prestige of scientists as high as it was in 1945. Indeed, perhaps May of that year saw its zenith; as Sir Michael Atiyah pointed out in his recent valedictory address as president of the Royal Society (see *Nature* 378, 525; 1995), the image of the benign, truth-seeking scientist suffered a damaging blow with the advent of nuclear weapons from which it has never fully recovered.

Few of the scientists who were engaged in the post-war reconstruction and are still alive today can be disappointed at the results of their efforts 50 years later. But many will no doubt still harbour regrets, some at the continuing failure of science to meet many of the world's needs, others at continued disputes about the future of nuclear weapons — the genie that is still not yet firmly back in the bottle — and new threats created by mankind's activities, such as those of global warming.

In the light of subsequent experience, many of the post-war ambitions turn out to have been highly optimistic. But without such optimism, the challenges of the future would remain daunting indeed. □

US science funding finally hits the frontier

Washington. Science is "like a fish swimming in water," says Norman Metzger, head of physical sciences at the National Research Council (NRC), the research arm of the US National Academy of Sciences in Washington DC. "It doesn't realize it's in water until you take the water away."

Since the end of the Second World War, the 'water' for US science has been the steady growth in funding — most of it from the federal government — which has supported new individuals and areas of research without directly threatening those already well established. But the tap has now been turned off; and 1995 was the year when the science community began fully to recognize that fact.

It is often said that Vannevar Bush, director of the Office of Scientific Research and Development set the stream in motion with his report, *Science — The Endless Frontier*, commissioned by President Franklin Roosevelt as a blueprint for the country's post-war scientific effort, and delivered to his successor, President Harry Truman, in July 1945.

The seminal report was confident in its proposals. A new National Research Foundation (NRF) should be set up, it said, with a budget growing rapidly to \$120 million a

year. It would, among other ambitious goals, provide 28,000 scholarships to boost a scientific workforce depleted by the demands of wartime military service.

Yet it is often forgotten that the central point of *Science — The Endless Frontier* was never implemented. The NRF — a true Department of Science proposal if ever there was one, and not entirely dissimilar to ideas currently being promoted by Republicans in Congress — took five years to get off the ground, in curtailed form, as the National Science Foundation (NSF). By that time the key areas of health and military research, which Bush expected to account for almost half of its budget, had been taken on by the new National Institutes of Health (NIH) and the Pentagon respectively.

Another myth about Bush is that his proposals were accepted without political rancour. In fact, his report was a clarion call for the support of science on scientists' own

terms, and was accepted in the face of opposition from Senator Harley Kilgore of West Virginia and others who argued that science should be directed to meet social and economic goals. That debate has never been fully resolved, and continues to this day: witness the recent efforts of Senator Barbara Mikulski (Democrat, Maryland) to foist "strategic goals" on the NSF.

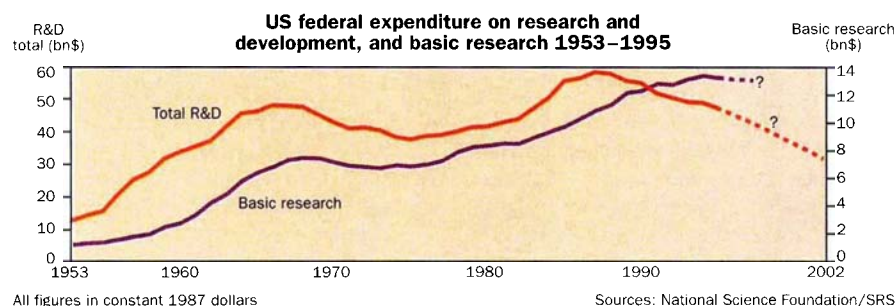
Yet it was not Bush's vision of science for its own sake, still less his failed proposal for an NRF, that drove the post-war explosion in federal support for research. Rather, it was the rapid expansion of mission-oriented research through government agencies. Sometimes the impetus came from the executive branch, such as the creation of the National Aeronautics and Space Administration (NASA), in October of 1958; and sometimes from Congress, as with the build-up of the NIH in the 1980s and early 1990s.

Money was available to satisfy this political pressure, as the annual budget spent at the discretion of the US Congress mushroomed with the US economy. Recession in the 1970s temporarily bucked the trend. But the overall trajectory was clear (see graph on next page). "Our central motif was: don't have a science policy," says Metzger. "One could do that in a time of fairly steady ►

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

**Bush: set framework
for US science in 1945.**

Taken from *Science — The Endless Frontier*



growth in budgets."

As a result, the story of US science policy in the past 50 years, suggests Metzger, is that of "the dog that didn't bark". Now it may have to. Total discretionary spending is due to fall sharply, as pensions and other burgeoning 'entitlements' eat up the federal budget. The only question is how successfully science and technology, which consume one-seventh of all discretionary spending, can deflect their share of the pain.

A National Research Council panel chaired by Frank Press last month published a report, *Allocating Federal Funds for Science and Technology*, which, if less awe-inspiring than *Science* — *The Endless Frontier*, is equally in tune with its times (see *Nature* 378, 426; 1995). The report asks the executive branch and the Congress to centralize control of the science budget so that difficult choices can be made — and adhered to.

But who will make such choices? Press did not — and says he was not asked to. Other scientists are equally reluctant; Martin Perl of Stanford University, winner of this year's Nobel prize in physics for his discovery of the tau lepton, told leading Republican congressmen at a meeting on 15 November that "old fields must die". But, he added quickly, "I'm not going to say which ones".

Perl pointed out that, before the days of extensive federal funding for science, participation in different disciplines would rise and fall quite rapidly. He reminded the congressmen of what many suspect — that the huge edifice of state support for science constructed since the Second World War incorporates a level of inertia that protects undeserving work, peer review or no. "In many fields, there are too many people and not enough good ideas," says Perl.

Admittedly, some choices are being made. Fusion research, fumbled by the administration, is being drastically curtailed by an impatient Congress. Bipartisan support continues to push more money into AIDS research and the Human Genome Project.

At the same time, environmental research is on a rollercoaster — sharply up two years ago, now crashing down again — for exactly the kind of reasons Vannevar Bush foresaw and sought, without success, to excise from the science policy process, namely that decisions on the optimal allocation of resources should be protected from short-term political interests.

Such choices will toughen as the millennium draws to a close. Last June, Al Teich, science policy analyst at the American Association for the Advancement of Science (AAAS), calculated that Republican plans

Germany: a phoenix tastes freedom

Munich. *Stunde Null* (Hour zero) marked the beginning of Germany's post-war reconstruction in all walks of life, including science. It is a term that describes the moment the Russians entered and occupied the Reichstag in May 1945; it was also a point at which Germany's entire infrastructure, including its research base, lay almost entirely in ruins.

Fifty years later, the success of the redevelopment of science in Germany is clearly reflected in the world-class quality of its research facilities and the steady production of Nobel prizewinners. Globally, Germany is the fifth largest spender on research and development, and the only country still increasing spending on basic research.

How did the phoenix rise from the ashes? In 1945, such an outcome was far from certain. Most of the universities and research institutes had been bombed almost out of existence. The allied forces had different ideas on how science should be rebuilt. Initially, the United States took the hard line that Germany should not be allowed to undertake any research at all.

Britain, in contrast, wanted to re-establish basic research as quickly as possible, and only to impose strict conditions on applied research. Its more liberal attitude appears to have been partly due to the influence of friendships between British and German scientists forged before the war, combined with a recognition that basic research was not seen as politically dangerous.

Many of the ideas about post-war reconstruction came from a group of leading German nuclear physicists, including Otto Hahn and Werner Heisenberg, both opponents of the Nazi regime, who had been captured by a secret US-British commando force and taken to Britain. Talks between these scientists, who were interned at Farm Hall near Cambridge, and British physicists, who had often been their pre-war colleagues, rapidly assured the Allies that Germany was far

to balance the budget will squeeze out one-third of the entire science and technology budget by the year 2002.

Now, President Bill Clinton has agreed to balance the budget too. When Charles Vest, president of the Massachusetts Institute of Technology, addressed a meeting in Washington last month, he joined a growing band of people who see the one-third cut not as a worst-case scenario but as a reasonable working assumption. Clinton's budget projections, Vest asserted, are not that much different from those of the Republicans. Fifty years after Vannevar Bush, the frontier is not as endless as it seemed, at least as far as science funding is concerned.

Colin Macilwain

from developing atomic weapons — and then concentrated on how science should be rebuilt.

One crucial decision, made at a meeting at the Royal Institution in London in October 1945, was that German research should be focused on the Kaiser Wilhelm Society (KWS), founded in 1911 and, before the war, responsible for 35 research institutes. This was to be renamed the Max Planck Society (MPS), and as such remains a major pillar of German research.

The change of name was approved in September 1946. But its acceptance beyond the British zone — which included Göttingen, to which most of the KWS's files and archives had been secretly taken by its war-time secretary, Ernst Telschow, in 1943 and 1944 — took a little longer. The United States resisted recognizing the society until persuaded to do so in 1948 by Otto Hahn, who had reluctantly agreed to become the first president of the MPS. A formal announcement of the founding of the newly named MPS, was made in *Nature* (161, 751; 1946), and by the summer of the following year, the organization was active in all three western zones.

Other ideas for post-war science policy in Germany had also been developed by the nuclear physicists during their internment. Heisenberg, for example, developed his idea for the Deutsche Forschungsrat (research

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Kaiser Wilhelm Institut für Chemie: renamed the Max Planck Society, it has underpinned German research.

council), which was set up in 1949, and in its short life was intended to function both as a funding agency and a central strategic advisory board for science. But this body failed to win widespread support, as it conflicted with the concern of German intellectuals, acutely sensitive to the dangers of concentration of power, that excessive federal influence over cultural matters should be avoided. It was transformed in the early 1950s into the Deutsche Forschungsgemeinschaft (DFG), today's funding agency for university research.

This wariness of centralized power remains the hallmark of German research. The German basic law, drawn up in 1949, states that "Art, science, research and teaching are free", and to ensure this, the *Länder* were given extensive powers in cultural affairs. But recognizing that it might not in practice be possible for research to have no central coordination at all, another article in the basic law opened the door to joint responsibility of federal and *Länder* governments for research.

But this door remained closed for years until two practical issues forced it to open: the inability of individual *Länder* to afford the large research facilities required to develop nuclear power, and the fact that poorer *Länder* were losing out to the richer *Länder* in terms of scientific development.

After the war, Germany was forbidden by the allies to undertake nuclear research. Indeed, the constitution precludes the establishment of a central atomic agency, as exists in most other countries. But in the 1950s Germany became as keen as others to develop cheap nuclear power, and the individual *Länder* could not afford to do this on their own.

As its first venture into the research arena, the federal government established a 'ministry for atomic questions' in 1955. Five national research centres concerned with nuclear research were rapidly set up. There are now 16 national research centres in Germany, and the political influence of the federal government is strongly felt, and resented, by scientists at the centres. They have come under attack in the past decade from the government which, as their paymaster, has demanded greater efficiency and closer collaboration with industry. The DFG and the MPS have more recently lent their support to the centres in their fight to ward off attempts to orientate their research to the needs of industry.

The issue of unequal development of research in the *Länder* has been addressed by a series of agreements from 1956 onwards, in which funding for research establishments is shared between federal and *Länder* governments according to agreed formulae.

The legacy of the war is still felt in the university sector. During the fascist years, the universities lost their traditional autonomy, and after the war they successfully fought to regain it. Federal influence is now

United Kingdom: promises, promises

London. In 1945, science in Britain was enjoying a peak of popularity that it has achieved neither before nor since. The British public had already been made aware of the contributions of scientists to a string of discoveries and inventions — from penicillin to radar — that had helped win the war.

At the same time, the intellectual freedom enjoyed by the research community was portrayed as a cultural ideal that contrasted sharply with the rigid ideology that had shackled Nazi researchers. "Never has science stood so high in public esteem" wrote *Nature* in May 1945.

This acclaim was reflected in the supreme self-confidence with which scientists faced the task of re-building Britain's — and Europe's — economic base. Those such as the crystallographer J. D. Bernal, whose calls in the 1930s for science to serve social ends had been viewed sceptically because of their links to socialist pro-grammes, now found their ideas had become part of the mainstream. As one historian puts it. "Having won the war, scientists were now going to win the peace".

At least four separate strands of the strategy for this debated at length during that period have made their mark over the intervening 50 years. Perhaps the most significant has been the political backing for links between science and production that has led to the growth of

Britain's science-based industries. The seeds of today's Technology Foresight programme for example can be seen in the debate over the need for strategic research priorities that took place in the columns of *Nature* and elsewhere.

A second, complementary, strand was the emphasis on the social value of basic research, and the need to respect the open communication of ideas. Britain had already had its own version of the Bush report (see page 5) in the work of the so-called Barlow committee of 1943.

A third element was the concept of planning for, through and by science. Debate at the time, reflected in *Nature* editorials, expressed a firm belief, rarely heard today, that rational planning was the key to the future health of all industrial economies. The planning failures of the 1960s and 1970s were still to come.

Perhaps the biggest gap between hopes and reality has been in the social status of science and scientists. "In the intellectual sense it is amazing what has been achieved in the past 50 years; but that increased knowledge has also led to a loss of sympathy, partly because of the failure of many scientists to explain their activities adequately," says Lord Flowers, former vice-chancellor of the University of London. Bernal, a great popularizer himself, would have shared Flowers' disappointment.

D. D.

felt only through support for university building and equipping, whose costs are shared, again to avoid disadvantaging the poorer *Länder*. But cracks in this system are also appearing, as the federal government now wants to cap funding to save money.

Strong decentralization has meant that Germany, rather than having a few central bodies controlling research strategy, has many forces whose views have to be juggled. Coordination at federal level is achieved through the science council, the Wissenschaftsrat, and the Bund-Länder Kommission (BLK), a political body that mediates between federal and *Länder* governments on cultural affairs.

This process has also recently been making itself felt in the former East Germany. The post-war development of science in the east was quite different from that in the west. Based on the Soviet model, research was removed from universities and placed in the centrally controlled institutes of the Academy of Sciences. Science suffered from tight political control, as well as the isolation from advances in the west.

Since reunification five years ago, western Germany has insisted on remodelling east German science in its own image, warts and all. The process of weeding out communists from the universities and research institutes was not as harsh as the denazification process overseen by the Allied forces immediately after the war (when, for example, 80 per cent of staff were sacked from the University of Munich, in the US sector). But the process has been painful, as has been getting used to a funding system based on scientific merit rather than political support.

One element — bureaucracy — that characterized both west and east has not disappeared. Decisions require consensus between all parties, including each *Land*. The decision-making process is therefore slow, cumbersome and compromise-orientated. But most Germans feel this process a small price to pay for the security that a pluralistic society provides. And most would also acknowledge that it has helped protect basic research against the attacks experienced in other industrialized nations.

Allison Abbott

Japan: from rags to riches, but where to go next?

Tokyo. In August 1945, when Japan's military finally surrendered, most of the country's urban centres lay in ruins, the economy was shattered and the populace was literally starving. The subsequent Allied occupation, which lasted until a peace treaty was signed in 1953, transformed the administrative, social and economic institutions — including universities and private and public research institutes.

The occupation, combined with General Douglas MacArthur's democratizing reforms, became the biggest single influence on Japanese science in the immediate post-war years. As in Germany, the occupying US Army, keen to end Japan's potential to wage war, banned research in aviation, radar and isolation of isotopes, even dumping the country's three cyclotrons in the ocean.

But other changes were more constructive. The university system was extensively reformed, resulting in the building of regional universities. Enrolments at universities increased 25-fold between 1945 and 1994, from 100,000 to 2,480,000, and the number of researchers in Japan increased sevenfold from 82,000 in 1960 to 558,000 in 1994, many of them engineers who helped to power the growth of Japan's industry.

Nevertheless, today, entrance to Tokyo University and the handful of other former imperial universities, as well as to a few of the older private universities such as Waseda and Keio, is still the principal goal of Japan's school-leavers.

Furthermore, with the student-age population in decline after peaking at 2 million in 1992, the gap between these elite old universities and the hundreds of new ones is likely to widen, at least in research, as the government plans to target more money at the best university research groups.

Up to the 1960s, post-war science focused on the importation and assimilation of foreign technologies, to aid economic reconstruction. By the mid-1950s, when Japan's gross national product (GNP) had been

restored to its pre-war levels, the government set up the Science and Technology Agency (STA), and initiated research programmes in nuclear power.

'Common use research centres' were created such as the Institute for Cosmic Ray Research at Tokyo University, and the Institute for Fundamental Physics, attached to Kyoto University. These centres at last provided Japanese academics with access to world-class equipment too expensive for individual universities.

This was followed in the 1970s and 1980s by the setting up of several institutes for joint university use, such as the Institute of Space and Astronautical Sciences (ISAS) which formerly belonged to Tokyo University. Several of these institutes, of which there are now more than a dozen, have achieved international recognition as centres of excellence, but they have tended to drain money away from research in the universities themselves.

Like much of the Japanese economy, science and technology are supervised and controlled by the national government. Academic staff at national universities, for example, are all direct employees of the Ministry of Education, Sport, Science and Culture (Monbusho).

Minoru Oda, former director general of ISAS, and also a past president of the Institute of Physical and Chemical Research (RIKEN), praises this close relationship. He says that while it is easy to criticize the government, the success of post-war Japanese science owes much to the direct support of government ministries.

Rapid growth of the Japanese economy from the 1960s onwards was matched by rapid increases in spending on research. Total spending on research was just ¥184.4 billion (US\$1.84 billion at current exchange rate) in 1960, only 1.1 per cent of GNP. But by 1990 this had risen to ¥12,100 billion, or 2.78 per cent of GNP.

But after five decades of economic growth achieved largely by adapting and improving technology introduced from overseas, Japanese science is headed for change that should lead to increased budgets for fundamental research.

The white paper on science and technology, published last year by the STA, states that Japan, having caught up with other advanced countries in the area of technology, no longer has models to follow and must now create new knowledge and new technologies of its own.

Recent supplementary budgets, for example, have contained funds

Revival for RIKEN

THE fortunes of RIKEN have, in a way, mirrored those of Japanese science. Severely damaged by bombing in the last year of the war, RIKEN was dealt a further blow when the Americans tossed its cyclotron into Tokyo Bay.

RIKEN, which had been a thriving and largely self-supporting centre for fundamental research since its foundation in 1917, was further weakened when it was stripped of its associated companies as part of general headquarters' policy of breaking up the *zaibatsu* or giant conglomerates which were seen as potentially dangerous to democracy because of the concentration of economic power that they represented. RIKEN struggled along for some years in a very poor state.

Help arrived, however, for RIKEN in the form of Kakue Tanaka, who became prime minister in 1972 and who as a young man had worked for the president of RIKEN. In a move supported by the many academics who remembered RIKEN's heyday, the government revived RIKEN with financial support, bringing it back to life in its present form as a largely government-supported research institute.

RIKEN is praised by one of its former presidents, Minoru Oda, as being "more academic than academia". Another special feature is its broad focus on all areas of science from fundamental basic research right through to industrial applications. Now, RIKEN is flourishing, with branches in Tsukuba Science City, Sendai and Nagoya as well as the main campus in northwest Tokyo. **S. B.**

for research grants available to researchers across the whole spectrum of public sector research.

Buffeted by recession, the effects of a strong currency and beset by challenges to Japan's technological lead in many areas of industry, policy-makers are betting that basic research can help to revive the faltering economy. This is in stark contrast to most other developed countries where the emphasis is now on 'strategic' research with clear socio-economic goals.

Indicative of this trend are the recently passed Basic Science and Technology Law, which guarantees government support for fundamental research, and a newly proposed law to establish a high-level committee to analyse and assess existing and proposed research projects.

If these developments indicate a genuine commitment to supporting innovative research, there can be no doubt that Japan, which rose so spectacularly from the ashes of defeat, will soon be making some extraordinary discoveries. **Stephen Barker**

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

France: Gaullist legacy casts a long shadow

Paris. France emerged from the Second World War with its infrastructure and laboratories ravaged by the German occupation. Fifty years later, France is a leader in many areas of basic science, and a high-technology showpiece. A nuclear weapons state, France has also dragged Europe into space, and provided the impetus for the development of Europe's Airbus airliners.

The foundations of the current system were laid immediately after the war under General Charles de Gaulle, and during de Gaulle's second term of office (1958–69). Science then stagnated under the presidencies of Georges Pompidou and Valéry Giscard d'Estaing (1969–81), and was rejuvenated under that of François Mitterrand (1981–95). Similarly, many of the constraints of the system originate in the strategic choices made during France's post-war reconstruction, and later during its efforts to re-establish itself as an international power.

The state began to play a significant role in French research in 1938, when it became the first country to appoint a science minister, Irène Joliot-Curie. The *Parti populaire* government proceeded, one year later, to create the Centre National de la Recherche Scientifique (CNRS), today Europe's largest fundamental research organization.

Both moves were aimed at closing the gap between industry and research, most of which was then carried out in the state-run universities. These refused to carry out applied research, while French industry lagged behind its competitors, who had developed large research-based companies in, for example, chemicals and electricity.

CNRS itself narrowly avoided being abolished in 1940 by the collaborationist Vichy regime. It emerged strengthened after the liberation, as many of its leaders were communists or had taken part in the resistance movement. But it was forced to modify its original goals during the late 1940s, as France began to seek alternative routes to harness research to the reconstruction of the country.

CNRS's original mandate to carry out targeted research was compromised when much of this was farmed out to a series of newly created agencies, such as the Commissariat à l'Énergie Atomique (CEA) and the agricultural research organization INRA¹. The existence of these new research agencies also prevented CNRS from fulfilling its mandate to coordinate a national research

policy. This had to wait until 1958, when de Gaulle returned to power.

CNRS nonetheless made a major contribution to the renovation of French science after the war. It provided a channel for money from the Rockefeller Foundation, for example, which provided much needed



The beginnings of the CEA: from left to right (seated) Pierre Auger, Irène et Frédéric Joliot, Francis Perrin, Lew Kowarski; standing, Bertrand Goldschmidt, Pierre Biquard, León Denieville and Jean Langevin.

equipment and supplies. More broadly, the creation of CNRS, together with the appointment of a research minister, initiated the institutionalization of state support for science. This trend gathered pace, after the liberation, when the state set up research organizations in a deliberate bid to catch up with the United States.

Indeed, the US model of carrying out research directed towards specific goals, which had proved so effective during the war, fitted well with the neo-Colbertist approach of de Gaulle, who chose detailed planning as the means to address the most pressing problems of reconstruction, such as developing energy, materials and transport. Where a strong base existed, as in nuclear power, electricity, and railways, France achieved rapid progress.

In fact, next to the reinstatement of the CNRS, the single most important event in post-war French science was perhaps the creation of CEA, which brought enormous resources to bear on the development of both nuclear energy and nuclear weapons, and provided broad stimulus to fundamental research in physics². (In contrast, French biology had to wait for its reconstruction until the late 1950s, when the government made molecular biology a priority, and spectacular advances were also made at the Institut Pasteur by André Lwoff, François Jacob and Jacques Monod³.)

The initial momentum created immediately after the war was lost, however, following the resignation of de Gaulle in 1946. Renewal began again in earnest only in the 1950s, first under the socialist government of Pierre Mendès-France — which organized a conference on research in 1956 at Caen, aimed at developing a national strategy for research — and in particular after 1958, when de Gaulle returned to power and established the Fifth Republic.

It was in that year, for example, that the government created an interministerial committee on research, as well as another widely acclaimed body, the Délégation générale à la recherche scientifique et technique. The role of the latter was to plan national research programmes in strategic areas such as molecular biology and space.

During de Gaulle's reign from 1958 to 1969, science spending increased from 2.45 to 6.2 per cent of the state budget, while the number of researchers jumped from 9,000 to 31,000. A wave of new agencies, including the French space agency CNES, were also created to reinforce the *grands programmes*.

The achievements resulting from de Gaulle's policies are undeniable. Electricité de France (EDF), for example, has established an efficient park of nuclear reactors, and there is now an impressive high-speed rail network. France has succeeded — where the United Kingdom and other countries have failed — in creating successful nationalized industries largely because these are run by élites, specially trained at the country's *grandes écoles*.

But state intervention — and these élites — is now coming under fire as ill-suited to both the increasing pace of technological change, and the need to adopt an aggressive strategy in a highly competitive global economy.

Indeed, many argue that the state's emphasis on the *grands programmes* has led to a neglect of innovation in the private sector, particularly in small companies⁴. Economic success also increasingly depends on sectors such as lasers, robotics, computing and biotechnology which are unsuited to state planning, and better served by fast-moving and flexible private companies.

Conversely, confusion about the role of CNRS may also have limited its effectiveness. From its beginning, the agency has been responsible for funding university research, and continues to fulfil that role today. But some observers argue that this dual role has made CNRS 'schizophrenic' by confusing its goals — namely support for its own research and that of the university system — and even inhibiting the development of a strong university system.

The only French universities carrying out international-level research are those that have large teams of CNRS researchers. This reflects the fact that, over the past 50 years, CNRS support has been a major factor allowing French universities to modernize themselves. But many now argue that the universities will be able to develop a strong research capacity only when this dependency is broken and they have regained full control of their research budgets. **Declan Butler**

1. Jean-François Picard, *La république des savants, La recherche française et le CNRS*, Flammarion, 1990.
2. Marie-José Lovérini, *Le Commissariat à l'Énergie Atomique*, Gallimard, 1995.
3. Michel Morange, *Histoire de la biologie moléculaire, La Découverte*, 1994.
4. Jean-Jacques Salomon, *La capacité d'innovation, in Entre l'état et le marché, L'économie française des années 1880 à nos jours*, eds M. Lévy-Leboyer and J.C. Casanova, Gallimard, 1991.

Unesco and population genetics

SIR — We are members of the United Nations Educational, Scientific and Cultural Organization (Unesco) International Bioethics Committee Subcommittee on Population Genetics but are writing in our personal capacity in response to a recent leading article and News article (*Nature* 377, 372–373; 1995). In some respects your articles may have misrepresented our final report, which was revised in response to the many constructive comments that we have received from population geneticists and representatives of indigenous groups. A copy can be downloaded from the Internet at: <http://www.biol.tsukuba.ac.jp/~macer/index.html>

The subject of the report is population genetics, of which the Human Genome Diversity Project (HGDP) is only one example, albeit the grandest in ambition and design. The report does not “strongly criticize the HGDP”, as implied by Declan Butler in your News pages. What we did criticize was the planners’ failure to anticipate the reactions of some groups of indigenous peoples to the HGDP, as well as the delay between the planning of the project and the release of detailed measures to explain how “informed” consent would be obtained. However, obtaining informed consent is an ethical problem for many population research projects. The latest measures produced by the HGDP planners contain some of the most ethically sophisticated and detailed procedures for obtaining informed consent from individuals and groups in population genetics research.

Many of the other issues raised in the report apply to the whole of population genetics research. One is commercialization and use of the results of the collected DNA and cells. This controversial issue is discussed in our report, to the extent that it should be addressed by those involved in the collecting of samples and be part of the consent and cooperation agreements. The resolution of biotechnology patents and “genetic prospecting” was not a subject for our subcommittee to resolve, but we urge the clarification of this issue by a global body.

One issue dismissed by the leading article in *Nature* is that of eugenics and racism. Although it is probable that little population genetic diversity will be found that is unique to one particular group, there is a logical possibility that there may be distinct genetic features that make one genetic group distinct from others.

While eugenics (cf. Galton) was founded on racism, eugenics today does not have to be linked with racism. Those who continue to link their eugenics with racism will not be dissuaded by scientific evidence, since racism is an attitude of mind, or prejudice. People who tenaciously hold to such

prejudices are usually not susceptible to the voice of pure reason and may even misappropriate scientific data to “prove” the truth of their own prejudices.

The fact that our report did not endorse any particular population genetics project is not a criticism of the HGDP. It rather is our personal view that the role given to us by Unesco could better be served, as could science, by the establishment of a separate ethical committee available to all population genetics researchers. Rather than being exclusively linked with one particular project and focusing attention solely on it, we want a wider selection of scientists to feel able to seek an international ethical committee, and we invite scientists to contact us.

Darryl Macer

*Institute of Biological Sciences,
University of Tsukuba,
Tsukuba Science City,
Ibaraki 305, Japan*

John Fleming

*Southern Cross Bioethics Institute,
PO Box 206, Plympton,
South Australia 5038, Australia*

Genoveva Keyeux

*Instituto de Genetica Humana,
Pontificia Universidad Javieriana,
Carrera 7a. No 40-62, Apartado Aereo,
56710 Santa Fe de Bogota, DC, Columbia*

Bartha Maria Knoppers

*Faculty of Law,
Université de Montreal,
CP 6128, Montreal,
Quebec H3C 3J7, Canada*

SIR — Your leading article “More focused agenda for Unesco” (*Nature* 378, 423–424; 1995) made two references to the International Council of Scientific Unions (ICSU) that merit correction. First, you state that “Unesco has been the chief source of support for ICSU”. Although our links with Unesco have been very strong since that organization’s foundation 50 years ago, Unesco’s yearly financial contribution to ICSU now represents a little over 10 per cent of our income, and this is used solely to support the scientific activities of our members. Second, you say that “Unesco was one of the prime movers in launching the International Geophysical Year”. In fact ICSU was the prime mover and sole sponsor of the International Geophysical Year which was a turning point in moving towards global science.

Unesco is, however, the sole intergovernmental organization devoted specifically to promoting science and technology, and it is a strong partner to the nongovernmental ICSU in a variety of scientific programmes in which both North and South participate. While the United Kingdom and the United States are full and active members of ICSU, their joining

Unesco would bring a great deal to that organization as well as to international science in general.

Julia Marton-Lefèvre

*(Executive Director)
International Council
of Scientific Unions,
51 Bd de Montmorency,
Paris 75016, France*

ORI defended

SIR — Mr Burton Appleton (*Nature* 378, 432; 1995) seems to have read only selected parts of the report of the ORI/AAAS Conference on Plagiarism and Theft of Ideas, the focus of his letter. He suggests there was “tampering” with the meeting transcript, but in the “Editor’s Note” on the second page of the report, I clearly stated on behalf of the Office of Research Integrity (ORI): “As an organizer of this Conference and editor of this summary report, I edited the talks and discussion for brevity and consistency, as well as to remove some comments that were *irrelevant* to the conference topic... or were relevant to *ongoing, non-public* investigations of scientific misconduct” [emphasis added], such as the active case he cites.

Appleton also claims that comment from him in which he “took ORI to task for notifying the public only through announcements in the *Federal Register*”, had been “struck from the record”. I did so because his statement was inaccurate, given the response by the director of ORI, Dr Lyle Bivens, to the first comments. As Bivens stated: “ORI policy now is that when a case is closed and there is a misconduct finding, we will publicize that through a notice in the *Federal Register* and in the *NIH Guide to Grants and Contracts*... The first time we published our *ORI Newsletter* that had the names of individuals found to have committed scientific misconduct in it [was in 1993]... The *NIH Guide* goes to some 34,000 addresses, and that is a very wide distribution in the scientific and academic community. The *ORI Newsletter* of course, goes to over 3,000 institutions that file assurances with our office, as well as to a substantial number of other people on the list. I think we give it very broad publicity.” The information in these three ORI notices and other publications is often picked up by the press as well. There is no “burying” of ORI’s findings as Appleton implies; ORI makes its findings very public.

ORI scientists are committed to upholding the principles of scientific integrity and to supporting open discussion of these issues, as we did when we sponsored this public conference on plagiarism.

Alan R. Price

*Division of Research Investigations,
Office of Research Integrity,
5515 Security Lane, Suite 700,
Rockville, Maryland 20852, USA*

Alphabetical disorder

SIR — In the article "China still hopeful ten years on", you mentioned¹ "... the sheer difficulty of learning the Chinese ideographs, which necessarily occupies several years of early schooling, raising the question whether China can successfully modernize itself without devising a more manageable script". This appears to betray a slight alphabetocentric tendency. One can equally say that it takes years of early schooling to learn the "unmanageable" spelling of English.

It is indeed true that learning Chinese ideographs takes much more time than learning the alphabet. Nevertheless, just to know the basic units of an alphabet does not convey anything about the language. Knowing the Roman alphabet tells one nothing about English, for example. Knowing the ideographs, on the other hand, actually tells one quite a lot about the language already, because each unit is a morpheme

in Chinese, while in the Roman alphabet each unit is at best a phoneme. This has been extensively discussed elsewhere²

It probably takes as much time to learn Chinese as to learn English. The difference is in the amount of time spent on learning different aspects of the written language. In Chinese it is the ideographs, while in English it is the way alphabetical units are strung together to make words. I have some limited experience of learning Babylonian, where the script, the grammatical structures and the syntactical structures are very different from either English or Chinese. A lot of time is spent on learning the *stati* of the nouns and the different forms (Gt, Gtn and so on) of the verbs, but this can be mastered without undue difficulty.

Moreover, the Chinese script co-evolved with the Chinese language over 30 centuries, and they are best suited to each other. Not many languages, with the possible exceptions of Sumerian (cuneiform logograms) and Korean (the Han'gul script), can lay claim to such an efficient system. In English, the spelling is so irregular and unpredictable that even native speakers such as Noah Webster or Bernard Shaw wanted to reform it or to adopt a new alphabetical system³. Part of the reason is that the Roman alphabet was an adaptation of the Phoenician alphabet⁴. The latter was designed for

Semitic languages which had few vowels. The Roman alphabet inherited this scarcity of vowels, and when it is used to write English, we see the problems that we are familiar with today.

You also mentioned¹ that "there is no real likelihood that China will abandon Chinese for some other language (English for example); the cultural tradition is simply too vivid and too strong". Why should we Chinese abandon our mother tongue for an alien language? Chinese may appear difficult to the non-Chinese, but to us it is an elegant and expressive language, easy to learn, beautiful in its spoken and written forms, and free of declensions, conjugations and such "absurdities" (when Chinese children learn an Indo-European language, they all have great difficulties with declensions and conjugations because they do not exist in Chinese). From a Western viewpoint Chinese may seem an inefficient language, but from a Chinese point of view all Indo-European languages appear cumbersome and unwieldy.

P.-L. Chau

Department of Pharmacology,
University of Cambridge,
Cambridge CB2 1QJ, UK

1. *Nature* **378**, 537–538 (1995).
2. Coulmas, F. *The Writing Systems of the World* (Blackwells, Oxford, 1989).
3. Haas, W. *Alphabets for English* (Manchester Univ. Press, 1969).
4. Gelb, I. J. *A Study of Writing* (Univ. Chicago Press, 1963).

Correspondence

Letters submitted for Correspondence should be typed, double-spaced, on one side of the paper only, or e-mailed to nature@nature.com

What are you doing in the office? Shouldn't you be in the lab running the monoclonal separations?

Evaluating data and monitoring runs throughout your company has been made easier for you. Much easier and more productive, in fact.

With UNICORN™ 2.00 automated control software, you can go from evaluating data to monitoring the most demanding runs as they take place. And you can do it without leaving your desk.

No matter where your Pharmacia Biotech systems are located, UNICORN 2.00 gives you access to them all. So you can control all systems from your computer—or

from any other PC in your company's network. What's more, you'll increase the security of your data as it can all be stored in one place, easily accessible for everyone who is authorized.

UNICORN 2.00 is the latest automated control software from Pharmacia Biotech. Just call us at 1 (800) 526 3593 in the United States, or +46 18 16 5011 from the rest of the world, for more information about UNICORN 2.00. That single phone call could be one of the most productive things you do this year.

Biosafety regulations

SIR — David Dickson's article (*Nature* 377, 94; 1995) on international regulatory agreements for genetically engineered organisms needs to be placed in context.

The international biosafety regulations being promoted by environmental regulators in the Netherlands, the United Kingdom and the European Union have some serious disadvantages. They would require case-by-case risk analysis by governments of field trials with genetically manipulated organisms (GMOs) that are largely of minimal intrinsic risk. The regulations would require new bureaucracies, as well as vastly increased costs to both governments and researchers. Under this scheme, organisms that are phenotypically identical to one another would be subject to significantly different regulatory requirements, determined solely by the method of genetic modification used.

The presumption that GMOs as a class of organisms are inherently dangerous has been exhaustively addressed — and discredited. The scope of GMO-specific approaches is incompatible with the broad scientific consensus, cited in an earlier leading article in *Nature* (356, 1–2;

1992), that “no conceptual distinction exists between genetic modification of plants and microorganisms by classical methods or by molecular techniques that modify DNA and transfer genes”. The OECD's Group of National Experts on Biotechnology, the US National Academy of Sciences and its research arm, the National Research Council, as well as other groups, have reached similar conclusions.

Beyond wasting human and financial resources, misguided approaches to regulation may increase the vulnerability of developing countries. For example, while the approach of the European Union and the United Nations Environment Programme focuses exclusively on organisms manipulated by molecular techniques, non-indigenous organisms (“exotics”) that are not GMOs are exempt. Introducing into a domestic agricultural ecosystem new organisms that have evolved in completely different ecosystems may well pose a greater risk than introducing organisms created through genetic modification of existing domestic germplasm.

Rational approaches to the testing and use of new agricultural products by developing nations is critical for addressing two kinds of danger. First, regulatory disincentives can threaten continued increases in local food supplies by dis-

couraging research and innovation in agricultural biotechnology. Second, the unregulated introduction of non-indigenous organisms poses a genuine and serious threat to the diversity of developing countries. What is needed are scientific and risk-based regulatory policies that will protect the environment and encourage, not impede, sustainable development.

Henry I. Miller

*Hoover Institution,
Institute for International Studies,
Stanford University,
Stanford, California 94305-6010, USA*

ums

SIR —

My keyboard's **plus** or **minus** sign

Can serve for simple sums,

but I've no **micron** sign on mine:

I'm substituting **ums**.

But when I publish **SEMs** or **TEMs**,

Maybe in years to come,

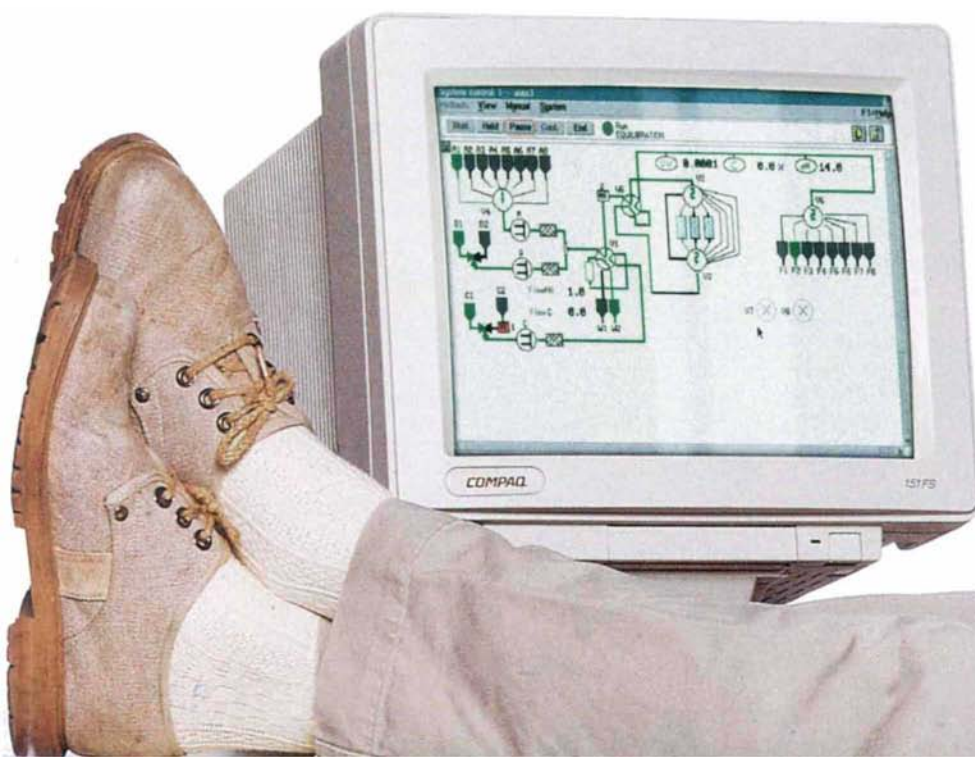
I plan to use a proper **mu**

Instead of printing **um**.

Ralph A. Lewin

*Scripps Institution of Oceanography,
9500 Gilman Drive,
La Jolla,
California 92093-0202, USA*

I'm doing them
right here with UNICORN 2.00



0 brave new words...

SIR — Recent work on the development of a UK licensing procedure for the use of bioremediation products in marine ecosystems shows that the term 'bioremediation' has been adopted by many groups to describe and promote their work and products. Conventional meanings have been swept aside. The time has come to establish an acceptable definition for the term.

The following are examples of the range of definitions in common usage in the literature:

"Bioremediation is the act of adding materials to stimulate the natural rate of biodegradation."¹

"Bioremediation is the application of biological process principles to the treatment of groundwater, soil, and sludges contaminated with hazardous chemicals."²

"Bioremediation is a managed or spontaneous process in which biological, especially microbiological, catalysis acts on pollutant compounds, thereby remedying or eliminating environmental contamination."³

These different ideas lead to confusion about the action of bioremediation agents. We have been approached by companies selling nutrients, bacterial preparations, surfactants and sorbents, all purporting to be bioremediation products. The UK licensing procedure will require a clear definition of bioremediation products, as this will determine both which products require testing, and the nature of the efficacy and toxicity tests that need to be applied. The crucial point of the science behind bioremediation has been the addition of materials to stimulate the natural biodegradative process (including co-metabolism). The added materials, whether nutrients, microorganisms or surfactants, are therefore the bioremediation products. Materials that do not stimulate the natural biodegradative processes should be called something different. As bioremediation agents can stimulate co-metabolic processes, we propose the following definition:

"Bioremediation is the act of stimulating the metabolism or co-metabolism of contaminants."

Note that the term 'intrinsic bioremediation', which has been used to mean "the study of the natural biodegradation rate"⁴, is wholly inaccurate, as nothing is done to stimulate contaminant biodegradation; it

merely involves monitoring natural processes. We suggest 'intrinsic biodegradation' as a more appropriate label.

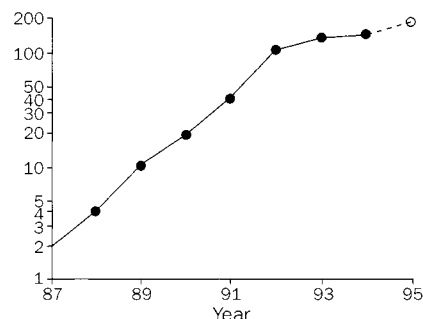
R. P. J. Swannell

D. J. Mitchell

M. A. Engelhardt

AEA Technology,
353 Harwell,
Didcot, Oxon OX11 0RA, UK

SIR — Since the term 'biodiversity' first entered the scientific literature in 1987 its use has continued to increase. As P. J. Hogarth has pointed out (*Nature* **364**, 664;



Number of papers with 'biodiversity' in the title. The 1995 total is extrapolated from data to October.

1993), until 1992 this rise was exponential, with a mean annual growth rate of more than 2. But a search of the BIDS ISI citation index shows that the annual rate of increase has slowed considerably since 1992 (see figure).

These data indicate that there will probably be an equilibrium density of around 250 papers a year with 'biodiversity' in the title compared with nearly 400 a year for '*Arabidopsis*' or more than 1,000 a year for '*Drosophila*'. Does biodiversity deserve even more attention from the scientific community?

Calvin Dytham

Department of Biology,
University of York,
York YO1 5YU, UK

Party line?

SIR — Your article on a recent speech by Sir Michael Atiyah, the president of the Royal Society, refers to a Labour Party science document that comments on the election of women fellows (*Nature* **378**, 525; 1995).

I should like to point out that the document referred to is not Labour Party policy. It was in fact written by the independent Science and Engineering Technology Forum, and was submitted to Labour as part of our Science 2000 consultation exercise, a focus for a long-term consultation process that Labour has been conducting with members of the science community. It is one of a number of

documents received, but the views and proposals it expresses are completely independent of myself and the Labour Party.

We shall publish our conclusions in due course at the end of the consultation process. Any policy that emerges will go through the usual policy-making channels of the Labour Party.

Adam Ingram

(Shadow Minister for
Science and Technology)
House of Commons,
London SW1A 0AA, UK

□ The phrase "a Labour Party document" was contained in the president's speech. The document in question was the only one presented at the launch of the Labour Party's Science 2000 exercise. — Editor, *Nature*.

Blood products

SIR — Your News article about my involvement in the debate about the safety of Armour's blood products¹ needs to be corrected. The article states that Armour continued to market Factorate "despite research showing that not all traces of HIV were destroyed by heat treatment". It quotes Corey Dubin as attacking my integrity more directly by stating that "[t]hey had virus [in the product] and Prince knew it; that is not a grey area". Both statements are untrue. In my letter to *The Lancet* published in early 1986² I expressed concern about the limited efficacy of dry heat inactivation, particularly when carried out for 10 hours at 60 °C. But I stressed that "this finding does not mean that dry-heat treated products are unsafe" as purification, processing and lyophilization steps may have efficiently removed or inactivated HIV.

I have been led to believe that Armour relied heavily on having been granted a licence for its product by the Food and Drug Administration (FDA), which in turn based its decision on the report from the Centers for Disease Control indicating that the Armour process was highly efficacious (~20 logs) in inactivating HIV³. Our work, whose results were reported to Armour, showed this report to be methodologically flawed.

Dubin is quoted as saying I should have communicated my results immediately to the FDA. But I believed that Armour was responsible for doing this. Lastly, Harold Sox's conclusion deserves to be emphasized; although companies hope to make a profit, none is so venal as to distribute HIV-contaminated products *knowingly*.

Alfred M. Prince

Lindsley F. Kimball Research Institute
of the New York Blood Center,
310 East 67th Street,
New York, New York 10021, USA

1. Appel, A. *Nature* **378**, 9 (1995).

2. Prince, A. M. *Lancet* **i**, 1280-1281 (1986).

3. McDougal, J. S. et al. *J. clin. Invest.* **76**, 875-877 (1985).

1. *Bioremediation of Oil Spills* Background Pap. OTA-BP-O-70 (US Congress Office of Technology Assessment, Washington DC, 1991).

2. Cookson, J. T. *Bioremediation Engineering — Design and Application* (McGraw-Hill, New York, 1995).

3. Johnson, M. B. & Larkin, M. J. presented at 132nd Ordinary Meet. Soc. gen. Microbiol. Poster 9 (Aberdeen, 1995).

4. Prince, R. C. in *Symp. 48 Soc. gen. Microbiol.* 19-34 (Cambridge Univ. Press, Cambridge, 1992).

5. Hinchee, R. E. et al. *Intrinsic Bioremediation* (Battelle Memorial Institute, Columbus, 1995).

Eighteen-ninety-six and all that

J. L. Heilbron and W. F. Bynum

This year's commemorative platter includes ether anaesthesia, smallpox vaccination, radioactivity and several mathematical morsels, as well as an accidental death (Otto Lilienthal) and the mother of all births (the Universe).

THIS year we can offer as anniversary of the year an invention more original in its design and consequences, more necessary to the correct thinking and wellbeing of humankind, more fundamental, inspired — we do not scruple to say awesome — than any other creative expression known

heading dates AMC (*ab mundo condito*).

5900 AMC (1.0 centenary)

Cleanliness, it is said, is next to godliness. Hence, in this jubilee year, Wilhelm Ludwig Friedrich Krafft's discovery a hundred years ago that soap solutions cleanse by colloidal suspension of offending material may deserve a celebration — and certainly does if combined with a nod to the washing machine (invented 1846, 1.5¢). It is also said that the world is a stage. So as it spins to its close, it might want to commemorate the installation in 1896 of the first high-tech revolving stage, by Karl Lautenschläger, at the Residenztheater in Munich.

One of the greatest theatres of science, the Institut de France, was the showplace for the most extraordinary discovery of 1896. For several months, Henri Becquerel entertained his colleagues with weekly reports of the darkening of photographic plates through their protective wrappings. The agent of darkness was a beautiful crystal of a phosphorescent salt of uranium exposed to sunlight before being placed with the wrapped film in a closed drawer. The odd experimental design derived from a conjecture by Henri Poincaré that, because X-rays originated in a fluorescent spot on a gas discharge tube, something interesting and invisible might arise from a phosphorescing salt.

One week the Sun did not shine in Paris. Becquerel nevertheless unwrapped his film (the show must go on), expecting that the week's bulletin would run: "no phosphorescence, no darkening". But the film was as dark as before. He inferred that his specimen radiated spontaneously. 'Becquerel rays' thus made their appearance along with Röntgen rays (last year's anniversary of the year, see *Nature* 373, 11; 1995), bringing France abreast of Germany in the discovery of odd emanations. Soon France outshone its rival with the detection of 'lumière noire' by Gustave Le Bon and rays from most yellow-green phosphors by Arsène d'Arsonval. All rays but Becquerel's soon flickered out, however; his, when followed up by Marie and Pierre Curie, brought to light the phenomenon of radioactivity.

One element of the Becquerel rays, named 'beta' by their

discoverer, Ernest Rutherford, turned out to be electrons. So did the agent responsible for the production of spectral lines, according to the analysis of Hendrik Antoon Lorentz of the splitting of lines in a magnetic field accomplished by his younger colleague Pieter Zeeman. Zeeman had persevered in his effort to influence light magnetically under the inspiration of Michael Faraday, who, 50 years earlier (1845, almost 1.5¢) had rotated the plane of polarization of light in a magnetic field. Lorentz and Zeeman shared the second Nobel prize for physics, in 1902; Becquerel received his in 1903, with the Curies.

Another high flyer of the time was the American engineer, astronomer and physicist Samuel Pierpont Langley (died 1906, 0.9¢), who successfully tested his steam-driven airplane over the Potomac River — until it crashed 1.2 km after take-off. Langley walked away. He was luckier than his fellow aeronaut Otto Lilienthal, who died in 1896 from injuries sustained in a glider accident.

Lilienthal's death and other vital statistics might have been recorded on machines sold by the Tabulating Machine Company, founded in 1896 by Herman Hollerith, whose punch cards, and the devices to read them, had revolutionized the 1890 census in the United States. TMC's descendant, IBM, is still with us.

Is it worthy of remark that, a hundred years ago, Jacques Hadamard proved that, if x is a very large number, the number of primes less than x is around $x/\log x$, said to be "the most important result ever obtained in number theory"; that Svante Arrhenius discovered the principle of the greenhouse effect and thus explained the ice ages; that Charles Edouard

Anniversary of the year: creation of the Universe. From Johann Jacob Scheuchzer's *Physica scara* (1731–1733).

to us. It is the act of creation that, on the scientific yet reliable calculation of the learned James Ussher, Primate of Ireland, happened in the year 4004 BC on the evening of 22 October, when God, being at leisure, decided to make a world. This coming October we have the singular opportunity of celebrating the sixtieth centenary of the creation of the Universe. Thinkers stuck in the pre-postmodern style of thought can suppose that the party opens the millionth millennium since the Big Bang.

Our anniversaries are arranged as usual, beginning with centennials, sesqui-centennials and so on back to 4004 BC, and ending with an anticlimactical list of memorable discoveries made in our century 50, or 50 ± 25 , years ago. Also, as usual, ¢ signifies 'centennial'. But for the convenience of those who need it, we depart from our usual custom to supply

A Hollerith printing, listing and tabulating machine.

Guillaume, director of the International Bureau of Weights and Measures, invented invar, a metal that scarcely dilates with heat; that Franz Exner, of Vienna, described his new method of measuring atmospheric electricity in fine weather; that Roland von Eötvös, of Budapest, published his first definitive measurements of gravity that later assisted the invention of general relativity; and that the German company Carl Zeiss took up the suggestions of Horatio S. Greenough, an American in Paris, and marketed the first stereo-microscope?

We think we know that in 1896 the longest refracting telescope in the world, 21 metres in length, went into operation at Treptow near Berlin, outdoing Chicago's Yerkes Observatory (18 m) in "licking the Lick" (15 m), California's first major scientific instrument. The licked Lick thereupon published the first photographic atlas of the surface of the Moon.

The application of these instruments had what may now seem to many people the desirable consequence of having no dangerous technological spin-off. In this they differed from the X-ray tube, which in 1896 was first used in therapy as well as diagnosis. The dangers and benefits are represented perfectly by the rationale and performance of the first therapeutical intervention, by a Dr Freund, who, reasoning from the destruction of hair on skin irradiated by Röntgen rays, used them to depilate a hirsute girl, "with complete success". At the same time, a Russian company marketed a substitute for stone, wood and metal called uralite, which could be sawn and nailed and resisted heat, fire, frost and acids. It was made of asbestos pressed with other minerals into sheets; like the cosmetic application of X-rays, uralite was intended to improve, not shorten, the lives of its consumers.

Neither X-rays nor asbestos injured Emil Heinrich du Bois-Reymond, a fierce pioneering physiologist, author of the positivistic war cry, "*Ignoramus, Ignora-*

bimus"; Armand-Hippolyte-Louis Fizeau, the collaborator of Jean Bernard Léon Foucault, of the pendulum, in proving, long after everyone believed it, that light moves more slowly in water than in air; William Robert Grove, a barrister by trade, the inventor of a standard electric cell and one of the many legislators of the law of the conservation of energy; Friedrich August Kekulé, a student of architecture turned chemist, discoverer of the key to the structure of organic compounds, the benzene ring; and Alfred Bernhard Nobel, captain of industry, inventor of dynamite and smokeless powder, and champion of peace, science and literature "of an idealist tendency"; nevertheless, they all died, 100 years ago.

5850 AMC (1.5 centenary)

Eben Frost and Gilbert Abbott, patients of William Morton, a dentist, and John Collins Warren, a surgeon, had good reason, in 1846, to sing the praises of Charles Thomas Jackson, who had proposed the use of ether as an anaesthetic; which was altogether more effective than the method recommended 200 years earlier (1646, 3.5¢) by Marco Aurelio Severino, who chilled the site of surgery with snow or ice. The patients survived their operations; Morton and Jackson eventually died bitter and insane. Nevertheless, we should all sleep easier because of what one surgeon dubbed this 'Yankee dodge'.

Another discovery of high drama took place in the heavens, when Johann Gottfried Galle turned his telescope to the point specified by the calculations of Urbain Jean Joseph Le Verrier (who was here behind, but independent of, John Couch Adams), and found the planet Neptune. The business took place almost at telegraphic speed: Le Verrier published his elements on 31 August, Galle found their (and Adams's) planet on 24 September. For our money, however, the real romance in the sky that year was the detection by Macedonio Melloni of the warmth of moonlight.

There is little enough warmth even in the sunlight of the Faroe Islands, whence the Danish government despatched a young medical graduate, Peter Panum, in 1846. His investigation of a measles outbreak (6,000 cases with 106 deaths) among the islanders is a classic epidemiological account of the behaviour of herd diseases in isolated populations.

One of the best scientific

Smithsonian Institution: "Establishment for the increase and diffusion of knowledge among men".

investments of all times was made in 1826 (1.7¢). James Louis Macie Smithson, an insignificant English chemist, left to the United States a bequest of £100,000, "to found at Washington, under the name of the Smithsonian Institution, an Establishment for the increase and diffusion of knowledge among men". The nonplussed but grateful nation did as required and, on 3 December 1846, appointed Joseph Henry, a physicist from Princeton (founded 1746, 2.0¢), the first head of the institution. It now has a dozen museums and galleries, a zoological garden, an observatory, a printing press and lord knows what else.

The seed of another investment, which has not been so happy for mankind, was planted by Christian Friedrich Schönbein's discovery of guncotton. The improvement of what we now call conventional explosives went on apace, culminating in the inventions of the melancholic Nobel, who, as previously announced, died a hundred years ago (1896, 1.0¢).

Let us not omit inventions that tend toward comfort. All praise belongs to Robert W. Thomson, a British manufacturer, who gave the world more pleasure than all the courtesans of Europe by his invention of an inflatable rubber rim for wagon wheels; Thomas Hancock, in the same trade, obtained a patent for moulding rubber articles for other uses; Alexander Bain eased the lives of telegraph operators with punched-paper tape; and Wilhelm Eduard Weber gave his colleagues, in his profound *Electrodynamische Maassbestimmungen*, a way to extend the familiar Coulomb distance forces to electricity in motion, sparing them for two generations the *Kopfzerbrechung* needed to follow Faraday.

We should note some comings and goings. This year 150 years ago saw the births of Antonio Abetti, Italian astronomer notable for his determination of the positions of asteroids (see under 5750 AMC) and comets; Magnus Gustaf Mittag-Leffler, Swedish mathematician and publicist for mathematics, on whose account, it is unreliably reported, Nobel refused to establish a prize for mathematics; Karol Stanislaw Olszewski, Polish



Daguerreotype of the second operation in which ether was used as an anaesthetic (17 October 1846) — the photographer was too squeamish to take pictures during the first operation, the day before. Warren stands in the right foreground; Morton is at the head of the operating table, centre rear of the photograph, wearing a checked vest.

chemist, a co-inventor of the first successful process for liquefying air; Edward Charles Pickering, called to direct the Harvard Observatory in 1876 (1.2¢), a tireless photographer of stars, whose observations of the spectrum of 'hydrogen' (in actual fact helium) figured prominently in the favourable acceptance of Bohr's quantized atom; Ira Remsen, chemist and educator, an important instrument of the spread of the German research ethic to the United States, and, as a recorded talking head, the keynote speaker of the current controversial Smithsonian (founded 1846, 1.5¢) exhibition "Science in American Life"; Emil Gabriel Warburg, German experimental physicist, discoverer of magnetic hysteresis, and an important force in the politics and pedagogy of German science; and the American Medical Association, an organization whose concern for the health of patients has been equalled only by its regard for the welfare of its members.

Against these we must set the deaths of Benjamin Waterhouse, American exponent of vaccination (see under 5800 AMC) and equally ardent opponent of alcohol and tobacco; and of Friedrich Wilhelm Bessel, who, after completing his apprenticeship as an accountant at the age of 22, began, in 1806 (1.9¢), to fashion himself into a most exact astronomer. Among his more famous works were the study of mathematical forms in the theory of gravitational perturbations (Bessel functions) and the detection of stellar parallax, which, like Fizeau and Foucault's measurement of the speed of light in water (see under 5900 AMC), proved what everyone already believed, in this case that Copernicus was more nearly correct than Ptolemy.

5800 AMC (2.0 centenary)

We recommend that observance of the second centenary of the publication of Pierre Simon de Laplace's *Système du monde* be joined with celebrations of the anniversary of the year. Laplace makes it clear how the trick of creation was pulled, from the swirling of the primordial nebula down to the formation and shaping of the Earth. And we might in the interest of universalism include a notice of the birth of the never-sufficiently-to-be-praised Johann Christian Poggendorff, the least of whose services to science was editing the *Annalen der Physik* for more than 50 years. Poggendorff belongs with the *Système du monde* because of his creation of the bible of historians of the physical sciences, his immortal (it is still being published) *Biographisch-Literarisches Handwörterbuch zur Geschichte der exakten Wissenschaften*.

In this year Carl Friedrich Gauss was not born, but born again, by his discovery of the constructibility (by straight edge and compass) of the regular 17-gon, a point missed by all the geometers of Greece, and of everywhere else; for, on the strength of this performance, he decided to devote his

life to mathematics. Among others born (for the first time) in 1796 were Sadi Carnot, whose inspired and perplexing theory of the perfect engine later stimulated invention of the concept of entropy and the second law of thermodynamics; Baden Powell, an expert on optics, an important spokesman for British science and, through his son Robert, first Baron Baden-Powell, the grandfather of all boy scouts; and Lambert-Adolphe-Jacques Quetelet, the Belgian astronomer and statistician father of *l'homme moyen*.

Further to statistics, David Rittenhouse, pioneering US colonial astronomer and clockmaker, Johann Daniel Titius, inventor in 1766 (2.3¢) of the Titius-Bode law of

Pierre Simon de Laplace: exposed the trick of creation in his *Système du monde* (1796).

planetary distances, and Johan Carl Wilcke, omni-competent Swedish experimental natural philosopher, translator and editor of Franklin's *Letters on Electricity*, died.

This toll, which could be lengthened, should not depress friends of good health. The year 1796 saw the publication of Christoph Wilhelm Hufeland's *Makrobiotik, oder die Kunst, das menschliche Leben zu verlängern*, which recommends daily air baths, thus evaporating the advice given two years earlier by the physician Samuel Gottlieb von Vogel, who preferred sea baths; and Edward Jenner recommended the bucolic pleasure of cowpox as a protective against smallpox. Jenner's vaccination of James Phipps, on 14 May, was a triumph of folk wisdom placed at the service of preventive medicine. Good health was also a preoccupation of Benjamin Thompson, Count Rumford, an American tory turned Bavarian nobleman, who introduced the potato into central Europe, published recipes for hearty soup, invented the kitchen range, argued for the kinetic theory of heat and, in 1796, to encourage the mind as well as the body, endowed the Rumford prizes for scientific achievement at the Royal Society of London and the American Academy of Arts and Sciences.

5750 AMC (2.5 centenary)

There were a great many books on diverse subjects published in 1746. Among our

favourites are Denis Diderot's *Pensées philosophiques*, which says most of what can be said to prove the existence of God from nature; Antoine Deparcieux's *Essai sur les probabilités de la durée de la vie humaine*, which introduced the concept of the life expectancy of the newborn; Pierre Bouguer's *Traité du navire*, the theoretical foundation of naval architecture; and Johann Heinrich Winkler's *Die Stärke der elektrischen Kraft des Wassers in gläsernen Gefässen*, in which, as reported last year, the author described the effects of using his wife to short-circuit a Leyden jar.

While on the subject of fulminations, we should mention John Roebuck's invention of the lead-chamber method for manufacturing sulphuric acid; the publication, by Jean-Etienne Guettard, of a map of the volcanic and other geological features of France; and the death of Colin Maclaurin, Scottish mathematician (Maclaurin's series) and fiery expositor of Newton. Among the births of 1746 were those of Jacques-Alexandre-César Charles, experimental physicist and showman; Giuseppe Piazzi (died 1826, 1.7¢), founder of the Palermo observatory and discoverer of the first asteroid, which he named Ceres after the patron goddess of Sicily; Benjamin Rush, physician, chemist and educator in the British colony of Pennsylvania; and Jan Hendrik van Swinden, Dutch electrician, magnetizer and meteorologist.

5700 AMC (3.0 centenary)

Of the making of books there is no end. During 1696 were published Guillaume François Antoine de l'Hospital's *Analyse des infiniment petits*, the first textbook on the differential calculus, and William Whiston's romance, *A New Theory of the Earth*, to name but two of scientific interest. That same year, we think, John Ray published the first description of peppermint; Johann Bernoulli challenged the mathematicians of Europe with the problem of the brachistochrone, whose solution led to the calculus of variations; and Johann Lotting set up the first ever factory for the manufacture of thimbles.

The three-hundredth anniversary of the death of Jean Richer should not slip by unobserved. Richer was the agent sent by the Paris Academy of Sciences in the early 1670s to the island of Cayenne to make measurements useful to astronomy and geography. While his confrères remained in the relative comfort of Paris, Richer sweltered to measure the refraction of the Sun, the parallax of Mars and the length of a seconds pendulum. His results enabled astronomers confidently to correct their observations for refraction, confirmed that the distance from the Earth to the Sun was much larger than even Kepler had allowed and provided an essential ingredient to Newton's deduction of the size and shape of the Earth.

Mary Evans

5650 AMc (3.5 centenary)

All that we can say of AD 1646 is that it was pregnant with things to come, or, to be exact, just finishing being pregnant. We refer to the births of two giants of industrious learning: John Flamsteed, who would become the first astronomer royal of the United Kingdom, and Gottfried Wilhelm Leibniz, who would become the first philosopher, and a not mediocre mathematician, of his age.

5600 AMc (4.0 centenary)

The same must be said about René Descartes, born in 1596, a latter-day Aristotle and Apollonius combined, the inventor of an analytic geometry, an implausible metaphysics and a world full of whirlpools, who inspired (sometimes with horror) all the active philosophers who knew his ideas.

One of his ideas was to base all of physics (which then meant all natural science) on laws of motion of inert matter — an idea that, when reworked in Cambridge, helped to produce the Newtonian system of the world.

5550 AMc (4.5 centenary)

The best bet for a 4.5 centenary is the folio by Georg Agricola containing tracts on mining and metallurgy (forerunners of his great *De re metallica* of 1556 (4.4c)) as well as his *De natura fossilium*. Agricola, a doctor who practised in the mining district of Joachimsthal, could describe in detail the various chemical and mechanical operations of smelting, extracting, assaying and pumping. No mere observer, he put his learning to use investing in mine shares, which made him wealthy. By 'fossils' Agricola meant anything found in the ground, including ingredients of the remedies described in Valerius Cordus's *Pharmacorum conficiendorum ratio*, the first German pharmacopoeia, also published in 1546.

0 AMc (60 centenary)

We know that our choice of 4004 BC for the date of creation might be controversial. People with more imposing natural scientific credentials than Ussher had arrived at other numbers, such as Michael Mästlin, the teacher of Kepler, who insisted on 4075 BC, and Giambattista Riccioli, SJ, whose *Astronomiae reformatae* of 1665 gave 4184 BC as the best weighted average of the sages of the ages. To which we reply that Mästlin and Riccioli were wrong. On their reckoning, the millennium has already come, whereas any fool can see that the old dispensation reigneth still. We acknowledge the possibility that Ussher too made the world too old and that nothing will happen on 23 October next. But who can be sure?

Although we have pressed our researches further into the past than usual, we have found nothing of scientific interest of certifiable and appropriate date worthy of an

anniversary between the creation of the world and the fifteenth century of our era. Our remote ancestors were inexcusably and, from a modern point of view, unintelligibly lax in supplying the data necessary to celebrate their discoveries.

5950 AMc (0.5 centenary)

The matter is altogether different for our century. The same machinery that guards priority and issues rewards has documented so many memorable events that we can notice only a few of the more characteristic of their era. The application of wartime technology and war-like funding to physical science gave a character to 1946. It saw the adaptation of captured V-2 rockets to solar observation; the start-up of the synchro-cyclotron at Berkeley, finished with money from the Manhattan District (the early US atomic bomb project) according to a design conceived by Edwin M. McMillan during quiet time at Los Alamos; the creation of the ENIAC, the first all-purpose, all-electronic calculator, by John William Mauchly and John Presper Eckert; the turn-on of the first Soviet nuclear reactor, under the direction of Igor Vasilievich Kurchatov; and the rapid development of radio-astronomy, which, still in 1946, secured the discovery, through Martin Ryle, of the first radio galaxy.

We also offer a few discoveries and inventions that did not require large machines or big money, but nonetheless were to have important ramifications in the postwar era: Willard F. Libby's introduction of radioactive dating by carbon-14; George Rochester and C. C. Butler's discovery of the first of the strange particles among the products of cosmic rays; Arthur Burks, Herman Goldstine and John von Neumann's foundational theory of computers, their "Preliminary discussion of the logical design of an electronic computing instrument"; and Joshua Lederberg and Edward Tatum's discovery of sexual goings-on in the reproduction of bacteria, reported in *Nature* 158, 558 (1946).

The Nobel prize for physics went to Percy Bridgman for creating the field of high-pressure physics; that for chemistry to James B. Sumner, John H. Northrop and Wendell M. Stanley for the preparation and crystallization of pure enzymes; and that for medicine to Hermann Joseph Muller for producing mutants with X-rays. It was a sign of things to come — all the winners were American.

5950±25 AMc (0.25c and 0.75c)

Several candidates for anniversarial notice for 1971 are developments of inventions

and initiatives of 1946. In space exploration, a descendant of V-2 rocketry, the Soviets successfully docked three cosmonauts at the world's first space station; two American astronauts brought back a sack of rocks from the Moon; the American Mariner 9 went into orbit around Mars two weeks before the Soviet explorer Mars 2 and three weeks before Mars 3 landed on the planet to broadcast unintelligibly for 20 seconds before going dead. On the computer front, Intel introduced the first microprocessor and Texas Instruments the first 'pocket' calculator (weighing just over 1 kg); Niklaus Wirth developed Pascal; and direct-distance dialling came in between Europe and the United States.

The Nobel prize for physics in 1971 was awarded to Dennis Gabor, a native of Hungary attached to Imperial College, London, for his invention of holography; that for chemistry to Gerhard Herzberg, a Canadian, for work on the electronic structure of molecules; and that for physiology or medicine to Earl W. Sutherland Jr, of the United States, for discoveries about the action of hormones.

Canada has further cause to celebrate this year. In 1921, Frederick Grant Banting and his collaborators started their attack on diabetes, which brought the world insulin and Banting and one of his colleagues a Nobel prize; and John Augustus Larson, a medical student, invented the polygraph or lie-detecting machine. We have other contemporary examples of the mechanization of mind: Karel Capek's play *R.U.R.*

(*Rossum's Universal Robots*), which introduced the concept and word 'robot' (from the Czech for 'worker'), was first performed; and Hermann Rorschach invented his routinized psychological probe, the inkblot test.

The Nobel prize for physics for 1921 went to nobody (the somebody to whom it went retrospectively the following year, Albert Einstein, was apparently still immature in the eyes of the judges); that for chemistry went to the same party, that is, to nobody (it would go in 1922 to Rutherford's former collaborator Frederick Soddy); and that for physiology or medicine to, yes, nobody (it would fall into the coffers of the prize-awarding body, the Karolinska Institute). Another clean sweep. Nobody then won prizes like Americans. □

J. L. Heilbron is at the Office of History of Science and Technology, University of California, Berkeley, California 94720, USA; W. F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BE, UK.



Captured and adapted: V-2 rocket.

Life (and death) in a malignant tumour

Kenneth W. Kinzler and Bert Vogelstein

Lack of oxygen induces the death of cells at the centre of tumours. But cells with mutations in the *p53* tumour suppressor can survive such conditions, providing yet another lead in the study of this notorious gene.

OBSERVATIONS that connect previously disparate areas of research often prove to be milestones. Investigation of the *p53* gene, which encodes a so-called tumour-suppressor protein, has in the past connected cancer genetics with viral oncogenesis, chemical carcinogenesis, transcription factors, programmed cell death, the cell cycle and drug resistance. The newest connection is made on page 88 of this issue¹, where Graeber and colleagues show that cells lacking intact *p53* are resistant to killing by hypoxia — that is, lack of oxygen. The results have implications for our understanding of tumour evolution, and link vasculature and oxygenation to the function and mutation of tumour-suppressor genes.

Over the past few years, it has been shown that one central function of the *p53* protein is to bind to DNA and regulate genes (including one called *p21*) that control cell birth and cell death. This is in part how *p53* suppresses unchecked cell division and therefore tumour growth. Research is now continuing along three lines: genetic, involving the description of *p53* mutations in various cancerous states; biochemical, involving the structure of *p53* and identification of associated macromolecules; and physiological, involving the effects of *p53* expression on normal cells and tumours. Progress in these areas seems to be correlated with the relative ease and sophistication of the three disciplines. Thousands of mutations have been discovered through genetics, and dozens of properties defined biochemically. But only two functions have been identified at the cellular level. In summary:

- Genetic studies have revealed that mutations of *p53* occur in diverse tumour types, and are clustered in the central region of the protein required for binding to DNA.

- Biochemical work indicates that virtually all mutations of *p53* abolish its ability to bind specific DNA sequences and activate the expression of adjacent genes. For example, the cyclin-dependent-kinase (cdk)

inhibitor *p21^{WAF1/CIP1}* contains two *p53*-binding sites within its promoter; cdk's are key players in cell-cycle control. The *p53* protein also binds to numerous proteins, including some involved in DNA repair and transcription². The carboxy-terminal end of

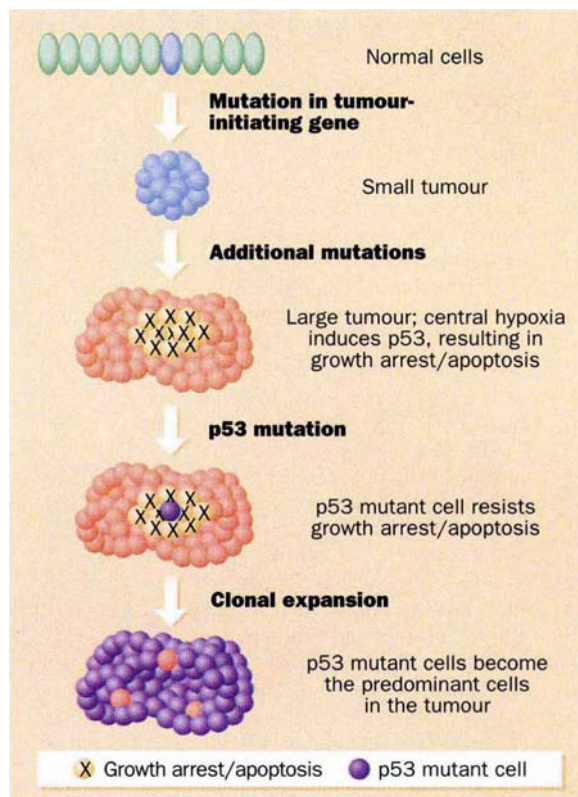
this regard concerns the conditions under which intact *p53* acts as a tumour suppressor. It is somewhat paradoxical that, unlike the products of other suppressor genes, *p53* is not expressed at detectable levels in the normal cells giving rise to common cancers. If it is not expressed, how can it normally exert tumour suppression?

The answer seems to be that *p53* is a sort of emergency brake, used only in stressful situations as a protective measure. Thus mice lacking *p53* are viable and healthy (except for a predisposition to develop tumours). But when the tissues of these mice are insulted, by gamma irradiation for example, their response is abnormal; cells do not undergo growth arrest or programmed cell death, either one of which can theoretically protect the organism from downstream effects of the DNA damage (principally uncontrolled growth and tumour development)⁷.

What is the stressful situation that normally initiates this process? Virtually all work on the topic has employed DNA damage, mediated by radiation or drugs, to activate *p53* (ref. 8). But tumours generally develop *p53* mutations in the absence of exposure to drugs or radiation, and it is likely that these experimental agents merely mimic some natural insult. A paper published last year suggested that hypoxia could also induce *p53* expression⁹. Graeber *et al.*¹ now confirm this result and relate it to events occurring during tumour formation. They

show that the low levels of oxygen found in those parts of tumours with poor blood supply result in *p53*-dependent programmed cell death. Mutations of *p53* reduce this cell death, giving the mutant cells a survival advantage over cells with intact *p53*.

All of this points to a plausible model for the involvement of *p53* in tumour evolution (see figure). Most tumours are initiated by genes other than *p53* and go through rounds of clonal evolution resulting in expansion. Angiogenesis — the formation of new blood vessels — is required for clinically apparent growth¹⁰; but, even with angiogenesis, the tumours eventually



Scheme for the involvement of *p53* in tumour evolution, incorporating the finding of Graeber *et al.*¹ that cells with mutant *p53* can survive the low levels of oxygen that kill other cells in the middle of tumours.

p53 mediates oligomerization of the protein and can promote the annealing of any two complementary strands of DNA³⁻⁵.

- From physiological studies it seems that the cellular consequences of these interactions are either growth arrest or programmed cell death (apoptosis). The effect observed depends on the cell type and the microenvironmental conditions studied⁶.

Although the techniques and approaches used in these studies vary, the physiology must be connected with the biochemistry and genetics if we are ever to understand *p53*, much less tumorigenesis in general. The first question that arises in

reach equilibrium with net cell birth equalling net cell death. Cell death is particularly prominent in the zones farthest from the tumour vasculature, where hypoxia stimulates apoptosis¹.

It is in this hypoxic witch's brew that p53 mutants arise. As in a scene from Dante's *Inferno*, these grotesque cells are born amidst the decaying remains of their ancestors, impervious to the noxious conditions which would destroy the merely mortal. They not only resist hypoxic death, but subvert the anti-angiogenic stimuli provided by cells expressing normal p53 (refs 11,12). They multiply uncontrollably, impervious to normal cell-cycle controls, and expand their mutant genomes through gene amplification and gross chromosomal abnormalities. It is only a matter of time before they escape from these grim surroundings, invade through adjoining virgin tissue and viciously attack innocent cells throughout the . . .

But wait! There seem to be some missing pieces here. What, for example, is the molecular messenger carrying the hypoxic signal to p53? Is the increased expression of p53 due to a direct effect of an altered redox environment^{13,14}, or an indirect effect mediated by the metabolic consequences of hypoxia, perhaps involving an unusual sort of DNA damage? Is the increased expression due to events during or after gene transcription? The p53 protein induced by DNA damage is stabilized, but the mechanism involved remains mysterious. And how does p53 induce apoptosis? It is now clear that activation of the gene encoding p21 protein is absolutely required for p53's ability to cause some human cells to arrest at the G₁ stage of the cell cycle¹⁵, that preceding the phase of DNA replication, and is necessary for at least part of the arrest in mouse cells^{16,17}. But the mechanisms underlying the ability of p53 to cause apoptosis are as yet

obscure. Deletion of p21 does not affect radiation-induced apoptosis in the mouse^{16,17}, and some (but not all) studies even indicate that the transcriptional activation properties of p53 are dispensable for apoptosis^{18–20}.

And the final connection, arguably the most important one, has not yet been made. Primary tumours usually do not kill cancer patients: they are excisable by the surgeon. Patients die because tumour cells spread from the primary site, replacing large parts of essential organs with cancerous tissue. The host of known oncogenes and suppressor genes, and their effects on signal transduction and

the cell cycle, explain how tumour cells subvert normal growth controls. But they do not explain the cellular migration and invasion of other tissues that by definition distinguish benign tumorous growths from cancers. We anxiously await studies that detail the biochemical pathways linking specific genetic alterations to those processes. Will this connection, too, involve p53? □

Kenneth W. Kinzler is in the Johns Hopkins Oncology Center, and Bert Vogelstein is in the Howard Hughes Medical Institute at Johns Hopkins, 424 North Bond Street, Baltimore, Maryland 21231, USA.

GALACTIC MAGNETIC FIELDS

Going with the flow?

Ellen Zweibel

MAGNETIC fields appear to be ubiquitous in the interstellar gas in disk galaxies. These fields have important effects on the dynamics and energy balance of the gas, as well as on star formation, and theory predicts that their morphology should reflect the structure of the ambient medium. However, the magnetic spiral arms discovered in the galaxy NGC6946 by Beck and Hoernes, reported on page 47 of this issue¹, do not coincide with any visible galactic feature.

The origin of galactic magnetic fields, and the processes that maintain them in their observed states, are still uncertain, but most of the current theories can be subsumed into one of two broad pictures. In one, the field reflects an approximate balance between resistive decay and amplification by rotation and turbulence; that is, the field is sustained by a dynamo. In the other, dynamo activity is long in the past, the present-day magnetic flux through any portion of the interstellar gas is virtually constant, and the field is said to be primordial.

Despite the many important differences between these pictures, both are based on magnetohydrodynamics (the science of magnetized conducting fluids), which dictates that magnetic fields are more or less 'frozen into' the interstellar gas and move along with it, somewhat as softening strands of spaghetti or coagulating ribbons of egg white are carried by vigorously boiling water. Thus, both pictures predict that magnetic structures and dynamical features should be correlated. To some extent, these predictions are confirmed by observation. The association of strong differential rotation (which stretches the field parallel to the flow) with the nearly azimuthal magnetic fields observed in galaxies is one such successful prediction. The observed alignment of fields with galactic spiral arms, although somewhat more complicated, is another.

In contrast, the magnetic arms in

NGC6946 seen by Beck and Hoernes do not correspond to any other observed structures. In polarized radio waves, the pattern of emission is symmetrical, with two main arms and two smaller ones, and extends at least over the optically visible disk of the galaxy (about 30,000 light-years), but the arms are only about 1,500 to 3,000 light-years wide (see Fig. 2 of Beck and Hoernes on page 48). The degree of linear polarization is consistent with a random field (which is small scale and therefore unresolved) equal to about half the large-scale field — an unusually high degree of order for a galactic magnetic field.

NGC6946 is a relatively nearby, gas-rich spiral galaxy with a high rate of star formation² (estimates of its distance range from 16 to 32 million light-years). A central bar may be present³, but if so it does not appear to drive the two-armed spiral structure theoretically predicted and sometimes seen⁴, for NGC6946 viewed at optical wavelengths has several arms. These arms are also seen kinematically⁵, in synchrotron radiation⁶, and in infrared emission from ionized carbon⁷. Although no velocity or intensity features in the gas distribution have been reported at the positions of the magnetic arms, the spatial resolution of the relevant studies was probably too low to have detected any such features if actually present.

What could be the cause of these narrow, extended magnetic arms? If they are

1. Graeber, T. G. *et al.* *Nature* **379**, 88–91 (1996).
2. Wang, X. W. *et al.* *Nature Genet.* **10**, 188–195 (1995).
3. Bakalkin, G. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **91**, 413–417 (1995).
4. Lee, S., Elenbaas, B., Levine, A. & Griffith, J. *Cell* **81**, 1013–1020 (1995).
5. Jayaraman, L. & Privess, C. *Cell* **81**, 1021–1029 (1995).
6. Haffner, R. & Oren, M. *Curr. Opin. Genet. Dev.* **5**, 84–90 (1995).
7. Cox, L. S. & Lane, D. P. *Bioessays* **17**, 501–508 (1995).
8. Kastan, M. B., Canman, C. E. & Leonard, C. J. *Cancer Metastasis Rev.* **14**, 3–15 (1995).
9. Graeber, T. G. *et al.* *Molec. cell. Biol.* **14**, 6264–6277 (1994).
10. Folkman, J. *Molec. Med.* **1**, 120–122 (1995).
11. Van Meir, E. G. *et al.* *Nature Genet.* **8**, 171–176 (1994).
12. Dameron, K. M., Volpert, O. V., Tainsky, M. A. & Bouck, N. *Science* **265**, 1582–1584 (1994).
13. Hupp, T. R., Meek, D. W., Midgley, C. A. & Lane, D. P. *Nucleic Acids Res.* **21**, 3167–3174 (1993).
14. Hainaut, P. & Milner, J. *Cancer Res.* **53**, 4469–4473 (1993).
15. Waldman, T., Kinzler, K. W. & Vogelstein, B. *Cancer Res.* **55**, 5187–5190 (1995).
16. Deng, C., Zhang, P., Harper, J. W., Elledge, S. J. & Leder, P. *Cell* **82**, 675–684 (1995).
17. Brugarolas, J. *et al.* *Nature* **377**, 552–557 (1995).
18. Caelles, C., Helmberg, A. & Karin, M. *Nature* **370**, 220–223 (1994).
19. Sabbatini, P., Lin, J., Levine, A. J. & White, E. *Genes Dev.* **9**, 2184–2192 (1995).
20. Haupt, Y., Rowan, S., Shaulian, E., Voudsen, K. H. & Oren, M. *Genes Dev.* **9**, 2170–2183 (1995).

1. Beck, R. & Hoernes, P. *Nature* **379**, 47–49 (1996).
2. DeGioia-Eastwood, K., Grasdalen, G. L., Strom, S. E. & Strom, K. M. *Astrophys. J.* **278**, 564–574 (1984).
3. Ishizuki, S. *et al.* *Astrophys. J.* **355**, 436–441 (1990).
4. Roberts, W. W., Huntley, J. M. & van Albada, G. D. *Astrophys. J.* **233**, 67–84 (1979).
5. Tacconi, L. J. & Young, J. S. *Astrophys. J.* **308**, 600–610 (1986).
6. Ehle, M. & Beck, R. *Astr. Astrophys.* **273**, 45–64 (1993).
7. Madden, S. C. *et al.* *Astrophys. J.* **407**, 579–587 (1993).
8. Cesarsky, C. J. *A. Rev. Astr. Astrophys.* **18**, 289–319 (1980).
9. Krause, M., Hummel, E. & Beck, R. *Astr. Astrophys.* **217**, 4–16 (1989).
10. Sukumar, S. & Allen, R. J. *Nature* **340**, 537–539 (1991).

formed by magnetohydrodynamic processes, then they almost certainly have a counterpart in the dynamics and/or density distribution of the ambient interstellar gas. The discovery of such features, or the establishment of meaningful upper limits on their strength, would undoubtedly furnish clues to the nature of the magnetic arms. It is natural to speculate that they may be associated with a central bar and/or with a possible past history of nuclear activity.

There remains the possibility that the arms are not magnetohydrodynamic in character. Synchrotron emissivity is, after all, a convolution of magnetic field strength with relativistic particle (or cosmic ray) intensity. The arms could represent an excess of cosmic ray electrons which illuminate the ambient magnetic field. If the excess is produced by a point source of accelerated particles (most likely in the nucleus of the galaxy, given the high degree of symmetry of the emission pattern), then the electrons must be able to propagate throughout the arms before losing their energy to synchrotron radiation. Taking representative values of 10^{10} electron volts for the electron energy and 5 microgauss for the magnetic field leads to characteristic radiative loss times of a few tens of millions of years. This far exceeds the light-travel time across the galactic disk, but cosmic rays in galaxies are thought to propagate by a mixture of diffusion and relatively slow advection⁸, which would lead to significant energy losses after several thousand light-years of travel. Thus, if the arms trace the path of relativistic electrons accelerated at a point source, the energy spectrum of the synchrotron radiation should 'age' with distance from the source in a predictable way, and could represent a test of this idea.

A spatially extended source of relativistic electrons would not display such spectral evolution. Such an extended source, however, should be accompanied by a magnetic feature, such as a shock wave or a current sheet (a thin region of magnetic reversal or otherwise strong magnetic shear), both of which are known to accelerate particles to high energies. In a sense, such a picture would be hybrid, relying both on an unknown magnetohydrodynamic process for forming such a structure and on plasma physics processes to accelerate the particles.

Finally, it is worth noting that similar magnetic arms have been detected in other gas-rich, actively star-forming galaxies^{9,10}, although not at the degree of spatial resolution reported by Beck and Hoernes. If narrow magnetic arms are found to be a common feature of galaxies, this itself must hold clues to their origin and nature, as well as making the problem of explaining them all the more compelling. □

Ellen Zweibel is in the Department of Astrophysical, Planetary and Atmospheric Sciences, University of Colorado, Boulder, Colorado 80309, USA.

Knowledge of the fount and the cause of disaster

William C. Evans

BEAUTIFUL and blue, Lake Nyos was about as popular a tourist attraction as its remote location in northwestern Cameroon would allow — that is until a huge eruption of carbon dioxide from the lake asphyxiated more than a thousand of the local inhabitants in August 1986

gas release, research there was limited to construction of a bathymetric map and collection of some samples of surface water.

In the weeks following the disaster, research teams from many countries visited the lake and collected a vast quantity



Lake Nyos four years after the disaster: a peaceful appearance hides the dual hazard of rising gas concentrations in bottom waters and a fragile natural dam being eroded towards the point of collapse.

(ref. 1). Occurring at night and leaving few nearby witnesses, the gas release presented several puzzles, including some impressive damage to vegetation (bending or uprooting) that was very asymmetrically distributed around the lake. The physical evidence, together with a small number of sketchy survivor accounts, indicated that the escaping gas did not bubble out randomly over the lake surface, but instead was mysteriously localized in an enormous fountain. On page 57 of this issue², Zhang helps us to visualize this fountain and develops the flow equations needed to assess possible fountain heights and exit velocities.

With a depth of 210 m and a surface area of $\sim 1.5 \text{ km}^2$, Lake Nyos is the deepest and one of the largest of Cameroon's 34 recognized crater lakes³. Before the

of data, but later at an international conference they failed to reach consensus on the mechanism of the gas release⁴. The localized fountaining was central to the disagreement, with some researchers arguing that it could result only from a high-velocity jet of volcanic gas from beneath the lake. Proponents of the opposing 'limnic eruption' hypothesis, which holds that the CO_2 released had been stored within the perennially stratified lake, argued for local uplift and decompression of gas-rich bottom water. The resulting exsolution of gas could then create a self-sustaining conduit within the water and ultimately produce fountaining high enough⁵ to account for the observed damage to vegetation 80 m above lake level¹. This controversy over the eruption mechanism has abated

1. Kling, G. W. *et al. Science* **236**, 169–175 (1987).

2. Zhang, Y. *Nature* **379**, 57–59 (1996).

3. Kling, G. W. *Limnol. Oceanogr.* **33**, 27–40 (1988).

4. Sigvaldason, G. E. J. *Volcan. geotherm. Res.* **39**, 97–107 (1989).

5. Tietze, K. in *Natural Hazards in West and Central Africa* (eds Freeth, S. J., Ofogbu, C. O. & Onuoha, K. M.) 97–107 (Vieweg, Braunschweig, 1992).

6. Nojiri, Y. *et al. Nature* **346**, 322–323 (1990).

7. Halbwachs, M., Grangeon, J., Sabroux, J.-C. & Villeville, A. C. *r. Acad. Sci., Paris* **316**, 483–489 (1993).

8. Kling, G. W., Evans, W. C., Tuttle, M. L. & Tanyileke, G. *Nature* **368**, 405–406 (1994).

9. Evans, W. C., Kling, G. W. & Tuttle, M. L. in *Water-Rock Interaction 8* (eds Kharaka, Y. K. & Chudakov, O. V.) 303–306 (Balkema, Rotterdam, 1995).

10. Mader H. M. *et al. Nature* **372**, 85–89 (1994).

somewhat with the discovery⁶ that gas concentrations in bottom waters are rising again after the release, and there is now general agreement that the dissolved gas should be artificially removed from the lake by the installation of pipes⁷. However, the long-missing mathematical model of an eruption fountain is a valuable new tool in assessing the hazards associated with gassy lakes, and any possible remedies.

Zhang argues that for a fountain to persist for any length of time and yet remain confined, the pressure profile inside the rising conduit must be near the hydrostatic gradient in the surrounding water column. The required flow conditions are thus different from those of fluid movement in pipes, where the walls allow large pressure imbalances. A fountain exit velocity can be calculated for any assumed CO₂ concentration (or corresponding saturation depth), and so Zhang's treatment can be applied directly to the general case of a gas-charged

lake where instability could develop at any depth below the chemocline⁸, the density boundary that prevents deeper seasonal flushing.

Regarding the 1986 Lake Nyos disaster, for a proposed maximum CO₂ concentration in bottom water of 0.43 mol kg⁻¹ (ref. 9), fountain heights in excess of 150 m could be reached. Decompressed to surface pressure, such fluid would theoretically have a porosity of 90 per cent, which is above the point of transition from bubbly flow to a stream of fragmented water droplets^{2,10}. During its most violent phase the fountain would thus resemble a scaled-up geyser more than an uncorked champagne bottle, the resultant mist in the moonlight possibly looking very much as one eye-witness described¹, like "smoke come out of lake". □

William C. Evans is at the US Geological Survey, Menlo Park, California 94025, USA.

NONLINEAR DYNAMICS

Avalanche theory in rice

Mehran Kardar

By patiently dropping grains of rice, in the experiment reported on page 49 of this issue, Frette *et al.*¹ put to the test some features of 'self-organized criticality' (SOC), a purported generic mechanism for the common occurrence of fractals, 1/f-noise and power-law distributed events. Although SOC is frequently observed in numerical simulations, controlled experiments on granular media have at best provided ambiguous results. Some piles of sand relax by big avalanches (of the order of the system size), whereas the statistics of avalanches in others can be fitted to different distributions. Depending on the shape of their rice grains, Frette *et al.* obtain both types of behaviour: small aspect ratios (short, fat grains) lead to large avalanches, larger ones (longer grains) lead to a power-law distribution for avalanche sizes — the signature of self-organized criticality.

A common exercise in elementary physics is to estimate the typical value of some quantity by dimensional analysis. Yet there exist many phenomena that are effectively scale-free, with possible values that span several orders of magnitude. Fractals — geometrical entities whose fluctuations repeat in self-similar fashion from the smallest to largest distances² — have been invoked to characterize natural shapes from coastlines to clouds. Possibly related manifestations of self-similarity in the time domain are the broad-band noise observed in electrical

circuits, and the fluctuations in time of the flow of rivers, or light from quasars. One model of this so-called 1/f-noise relies on the incoherent superposition of signals from oscillators with a wide (scale-free) distribution of frequencies *f*. Power-law distributions also seem to characterize the size (strength) of many natural catastrophes: avalanches, mass extinctions and stock-market crashes, to name but a few. The Gutenberg–Richter law, which relates the number of earthquakes to their magnitude, is a prominent example³.

Why are scale-free phenomena so frequent in nature? In 1987, Bak, Tang and Wiesenfeld⁴ introduced the idea of self-organized criticality as a possible answer. They used the analogy of a pile of sand gradually built up by the addition of single grains. Once the pile has reached a critical slope, addition of further grains may cause avalanches, while the overall pile fluctuates around this 'angle of repose'. In the idealized sand pile, modelled on the computer, avalanches of all sizes occur at this stage, with a probability distribution that is scale-free (a power law).

The initial claims of SOC as the underlying link between fractals, 1/f-noise and power-law distributed events naturally excited the imagination of a broad spectrum of scientists. Currently, the principle is applied more modestly to explain power-law distributions of events (avalanches) observed in some slowly

driven systems. Earthquakes, forest fires and rainfalls are just some of the diverse phenomena that have been suggested to show SOC.

As the original theory was cast in terms of avalanches on a sand pile, granular systems have been a prime target in the experimental search for SOC. But the behaviour of real sand is complex, showing a variety of slip, slide and roll motion not usually incorporated in the simple dissipative computer models. Some piles relax in large avalanches where a top layer of sand slides away⁵, and even though a spectrum of avalanche sizes is observed in other experiments^{6,7}, the limited range of data makes a convincing identification of SOC difficult.

Frette *et al.* set out to do the definitive experiment on this subject, and they have carried out the task with great patience and care. They do their experiment on grains of rice, which they find to be less subject to slipping and rolling than are sand grains, and hence closer to the dissipative limit studied numerically. (This is especially true for long-grain rice, with its larger aspect ratio.) They accumulate sufficient statistics to distinguish between a power-law distribution in one case (long grains) and a stretched exponential in another (more spherical grains). The occurrence of both behaviours in the same experiment is valuable, and allows the authors to provide possible reasons for distinct relaxation dynamics by appealing to the relative importance of inertia and friction. Although this explanation is not new, here it is clearly demonstrated.

Whether or not SOC is indeed a useful model (and the debate is not yet settled), it has certainly generated broad excitement. The new experiment provides the best evidence so far for the occurrence (sometimes, but not always) of power-law distributed avalanches in granular media. The continuing challenge to both theorists and experimentalists is to go beyond observation of power-law distributed events to the identification of the exponent and demonstration of its universality in different systems — that is, to go from qualitative to quantitative agreement. □

Mehran Kardar is in the Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

1. Frette, V. *et al.* *Nature* **379**, 49–52 (1996).
2. Mandelbrot, B. B. *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).
3. Gutenberg, B. & Richter, C. F. *Ann. Geofis.* **9**, 1 (1956).
4. Bak, P., Tang, C. & Wiesenfeld, K. *Phys. Rev. Lett.* **59**, 381–384 (1987).
5. Jaeger, H. M., Liu, C.-H. & Nagel, S. R. *Phys. Rev. Lett.* **62**, 40–43 (1989).
6. Held, G. A. *et al.* *Phys. Rev. Lett.* **65**, 1120–1123 (1990).
7. Bretz, M., Cunningham, J. B., Kurczynski, P. L. & Nori, F. *Phys. Rev. Lett.* **69**, 2431–2434 (1992).

The gut in feast and famine

Andrew R. Cossins and Neil Roberts

WITH the season of excess just over in many parts of the world, now is as good a time as any to ponder the physiological effects of overeating. The post-prandial slump is one such consequence, as blood is diverted to the intestines and liver, and the rest of the body relaxes to expedite digestion. Two papers by Secor and colleagues^{1,2} show that the demands we might put on the digestive system by eating especially large meals are nothing compared to those experienced by some snake species. These studies provide a new and spectacular model for investigating how the digestive physiology of vertebrates may be adapted to altered functional needs.

several months. But when a meal does arrive it can be of gigantic proportions. It is not unusual for pythons to consume up to 50 per cent of their own body weight at one sitting and there are published accounts of up to 160 per cent³. These enormous meals can be digested and assimilated in just a few days. Exactly how the snakes' digestive systems respond to extended periods of fasting followed by massive feeding was unknown.

Secor and colleagues have therefore followed the changes in intestinal structure and function in the Burmese python¹ and sidewinder rattlesnake² during the few days after feeding. They show that the small

gastric hydrochloric acid, and its oxygen partial pressure plunges to a quarter of its normal value, indicating the intensity of aerobic metabolism. Indeed, Secor and Diamond¹ show that whole-animal oxygen consumption of pythons increases 17 times only 24 hours after feeding (that of the rattlesnake increases seven times²). This huge factorial increase is comparable to that shown by sprinting mammals, but pythons maintain the increase for several days while remaining virtually motionless.

Of course, all of this increased activity comes at a considerable energetic cost. Secor and Diamond estimate that some 32 per cent of the metabolizable energy of the food in the python is consumed during the oxidative burst that supports digestion, against 10–23 per cent in animals that feed more-or-less continually. Although this extra energetic cost is attributable to mucosal upregulation, it is more than offset by the energetic savings from not maintaining a functional mucosa during the long period between meals.

These two studies provide an unusually clear picture of the remarkable extent to which physiology can respond to the fluctuating demands imposed by unusual lifestyles. They also open up some new research options in the study of gastrointestinal function. First, in contrast to virtually all other vertebrates, the digestive system of snakes is linear from mouth to anus⁴, and this morphological simplicity facilitates analysis. Second, the size of large snakes such as the Burmese python makes them convenient for study by whole-body magnetic resonance imaging (MRI). Conventional MRI can monitor gastric emptying and peristaltic movements⁵; high-resolution MRI techniques can characterize and monitor the response of the intestinal mucosa; and MR spectroscopy with phosphorus-31 offers the prospect of directly and non-invasively investigating metabolic processes in a snake's internal organs and tissues, including quantification of pH changes during digestion. It is a fair bet that many of the gastrointestinal processes in snakes are evident in other animals, albeit in less extreme form. □

Gunter Ziesler/Bruce Coleman

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Tough act to swallow — a rock python makes a meal of a Thomson's gazelle.

Digestion and assimilation consume a considerable fraction of the energy made available in the food itself. This energy is required not only to power digestion, absorption and the processing of food, but also for the maintenance of the digestive structures themselves. The mucosal lining of the vertebrate small intestine is a morphologically complex tissue with a very active metabolism. This, together with the remarkably high rate of cell loss and renewal, make it an energetically expensive tissue to maintain in prime condition³, and animals that are periodically under severe energetic constraints may not be able to sustain the cost. An example is that of ground squirrels, which undergo a noticeable reduction in mucosal thickness during hibernation⁴; hibernation is a strategy for energy saving and part of the strategy is to shut down expenditure on maintenance of the intestine.

A second case is that provided by those bradymetabolic animals (that is, having a low energy consumption) that feed intermittently. The most extreme examples are the vipers, boa constrictors and pythons, which, being sit-and-wait predators, eat infrequently, their meals often being separated by long and unpredictable fasts of

intestine of fasting individuals has an atrophied mucosal lining. Feeding, however, initiates a period of rapid reconstruction and the weight of the small intestine in both species doubles within a few hours entirely due to an increase in mucosal thickness. This massive reconstruction project occurs well before any assimilation of the meal can occur and the costs therefore must be met by committing energy reserves well before any payoff.

Measurements of nutrient absorption in isolated intestinal preparations show that the functional capacity of the system increases far more than expected from the doubling of the mucosal weight. Thus uptake rates for radiolabelled amino acids at the mucosal brush border in pythons increased by six times after 24 hours and by 16 times after three days, probably because of an increased capacity of the underlying transport systems combined with a doubling (or more) of mucosal surface area. Interestingly, in pythons these upregulated intestinal activities greatly exceed those of another bradymetabolic, continuously feeding carnivore (rainbow trout) and bear comparison with those of mammalian carnivores. The blood shows a strong alkalosis, probably due to the intense production of

Andrew R. Cossins is in the Department of Environmental and Evolutionary Biology, and Neil Roberts is in the Magnetic Resonance Centre, University of Liverpool, Liverpool L69 3BX, UK.

1. Secor, S. M. & Diamond, J. *J. exp. Biol.* **198**, 1313–1325 (1995).
2. Secor, S. M. & Nagy, K. A. *Ecology* **75**, 1600–1614 (1994).
3. Johnson, L. R. in *Physiology of the Gastrointestinal Tract* (ed. Johnson, J. R. Jr) 301–333 (Raven, New York, 1987).
4. Carey, H. V. & Sills, N. S. *Am. J. Physiol.* **263**, R517–R523 (1992).
5. Green, I. I. W. in *Biology of Pit Vipers* (eds Campbell, J. A. & Brodie, E. D. Jr) 107–111 (Selva, Tyler, Texas, 1992).
6. Parker, H. W. & Grandison, A. G. C. *Snakes—A Natural History* 2nd edn (British Museum of Natural History/Cornell Univ. Press, 1977).
7. Stehling, M. K. *et al. Radiology* **171**, 41–46 (1989).

The attraction of sand dunes

Robert S. Anderson

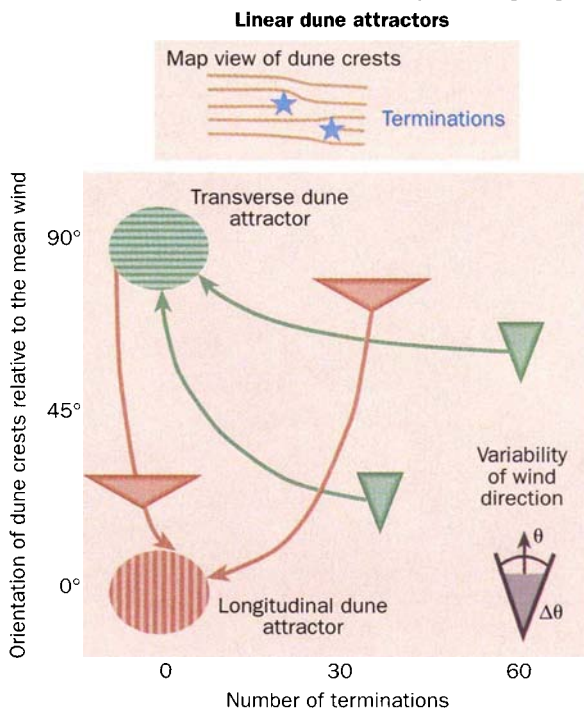
SAND is organized by wind into a broad spectrum of forms, the smallest being ripples of order 10 centimetres in wavelength, the largest being sand dunes up to a kilometre across. The simplest and surely the most photographed dunes are barchan dunes: isolated, crescent-shaped in map view, with arms or horns pointing downwind. Linear dunes are more space-filling, and can cover vast expanses of desert, dominating the huge sand seas. In map view they are characterized by long, nearly linear crests, and have terminations and bifurcations akin to those found in fingerprints. Curiously, they come in two varieties, with crests oriented either normal to the mean wind direction (transverse dunes) or parallel to the mean wind (longitudinal dunes). Then there are the star dunes, huge piles of sand with several arms swinging away from the main body. Geomorphologists have long sought explanations for this variety of forms and, in a paper published last month, B. T. Werner demonstrates a numerical model that successfully embodies the essential physics of the dune system without becoming ensnared in the details¹.

The fundamental processes acting on all dunes are the same. Dunes move downwind by a combination of two processes: saltation and grain flow. Sand bounces up the upwind side of a dune in the short, wind-aided hops characteristic of the saltation ('jumping') process, leading to the development of a depositional bump just to the lee of the crest of the dune, where the airflow diminishes. These bumps eventually oversteepen and fail, feeding sand into grain flows or sand avalanches which transport the sand down the lee face of the dune.

The problem has been to explain why dunes of distinctly different shapes evolve if the processes acting are always the same. The arguments have tended to revolve about the roles of the variability in wind direction, and the availability of sand for transport. It is generally believed, for instance, that barchan forms are associated with sparse sand supply and a unidirectional wind.

As sand dunes are composed of 10^{13} – 10^{20} sand grains, Werner's modelling strategy could not involve solving the equation $F = ma$ for each grain through its many trajectories, as was appropriate for the exploration of the fundamental physics of saltation², or of the evolution of smaller

forms such as sand ripples^{3,4}. Instead, Werner settled on a strategy in which slabs of sand are moved about on a lattice using simple transport rules that crudely mimic the processes involved. At every click of the clock, a slab of sand is picked up from a random site and transported in the direction of the instantaneous wind. If the sand sees a rough stony bed after travelling some fixed distance, it is more likely to keep



Phase space showing two distinct linear dune attractors. Schematic trajectories show how dunes would evolve towards one attractor or the other, depending on the degree of variability in the wind (depicted by the width of the triangles). All dunes in these wind regimes evolve from right to left in this phase space, towards cleaner patterns having fewer terminations, or defects within the dune field (stars in the map view). More unidirectional winds (green) lead to the formation of long-crested dunes with crests orientated normal to the prevailing wind (transverse dunes), whereas more variable winds (red) result in long-crested dunes aligned with the mean wind (longitudinal dunes).

going, whereas if it encounters a sand bed, it is more likely to be deposited. This takes account of Bagnold's observation⁵ that saltating sand bounces more efficiently off stony desert surfaces than off sand. The other main rule enforces the angle of repose: if deposition at a site results in a local slope of greater than 30°, the pile is adjusted by moving slabs downhill until no slopes violate the angle of repose.

The results are striking. Dunes inevitably evolve from an initially random topography, and the shapes that appear in the calculations closely resembled the various known dune patterns. In addition, typical dune dynamics are seen. For instance, barchan dunes calve off smaller dunes from their

horns; and terminations and bifurcations in linear dune fields act like dislocations in crystals that move downwind.

As intriguing as the dune calculations themselves is the overlay that Werner uses to address the reasons for the distinct shapes. Given today's computational power, it is relatively simple to set up and run models that embed abstracts of geomorphic processes. The effort lies in interpreting the results, pulling out global conclusions from a set of model runs in which a large array of parameters can be adjusted. Why did the system evolve to this pattern? How stable is the result, and how sensitive is the final pattern to the initial conditions? What sorts of quantities might we measure in the field that would add to our understanding at this dune scale?

Werner argues that at the scale of dunes, the system becomes independent of the microphysics of the processes involved, and obeys instead some over-arching rules. For instance, small dunes travel faster than large dunes, catching them and merging with them. Noting that this and other aspects of the model behaviour is akin to that of complex systems, Werner casts the problem of interpretation into the terminology of nonlinear systems. He defines a phase space within which the system evolves, and notes that there are several possible attractor states (preferred spots in the phase space toward which the system evolves, independent of the initial condition).

The art in this approach lies in defining the state variables. To address the linear dune problem, for instance, Werner chooses the variables 'number of terminations', and 'dune orientation relative to the mean wind' (see figure). He finds that the enigma of linear dune orientation can be explained by the degree of variability of the wind about the mean: large variability yields longitudinal dunes, small variability transverse dunes. Two spots in this phase space are attractors, the phase-space trajectories from any initial condition being dependent on the variability of the wind. This supports an earlier argument based on a small-scale experiment⁶, and seems to lay to rest the need to appeal to a longitudinal roller flow pattern in the atmospheric boundary layer. In all

1. Werner, B. T. *Geology* **23**, 1111–1114 (1995).
2. Haff, P. K. & Anderson, R. S. *Sedimentology* **40**, 175–189 (1993).
3. Landry, W. & Werner, B. T. *Physica D* **77**, 238–260 (1994).
4. Anderson, R. S. & Bunas, K. L. *Nature* **365**, 740–743 (1993).
5. Bagnold, R. A. *The Physics of Blown Sand and Desert Dunes* (Methuen, London, 1941).
6. Rubin, D. M. & Hunter, R. E. *Science* **237**, 276–278 (1987).
7. Werner, B. T. & Fink, T. M. *Science* **260**, 968–971 (1993).

cases, the dune pattern arises not by reflecting the pattern of the fluid flow itself, but rather is an inevitable result of the coupling between fluid flow and sand transport (a result mirroring conclusions from an earlier modelling effort to explain regularly spaced beach cusps⁷).

What might we learn from this approach? Although the terminology of phase space and attractors may be foreign to the general geomorphological and geological community, we have a great deal to gain from casting these complex problems into this arena. Many of the problems faced

by those interested in landscape evolution involve linked processes whose microphysics and chemistry tangle to create complicated landscapes. Werner's demonstration of the utility of this approach in the relatively simple case of sand dunes should inspire other geomorphologists to explore models of more complex landscapes for sets of simplifying rules. □

Robert S. Anderson is in the Department of Earth Sciences, University of California Santa Cruz, Santa Cruz, California 95064, USA.

MATERIALS SCIENCE

Getting oneself together

Philip Ball

"WHEN I use a word, it means just what I choose it to mean — neither more nor less." That pronouncement could surely come from the lips of materials scientists talking about self-organization and self-assembly (although its true originator, Humpty Dumpty, was no master of the latter). Substitute 'phase separation' or 'crystallization' and at a stroke your work sounds decades older. Nonetheless, striking examples of the spontaneous ordering and patterning of materials can be found that genuinely warrant such neologisms; and several were described at a recent meeting of the Materials Research Society*.

At the meeting, an entire symposium was devoted to self-assembled materials, which found itself both figuratively and literally next door to that on materials inspired by biology. Yet it was elsewhere that Jerry Tersoff (IBM) was to be found explaining how deceptively simple physics can be an organizational force too. Picture crystals of a germanium-silicon alloy growing on the surface of silicon. The surface energies are such that the crystals grow as tiny, faceted pyramids, the corners of a cubic phase. These are self-assembled quantum dots in some people's books, humble semiconductor crystallites in others. Their growth follows random heterogeneous nucleation events, so that the silicon surface becomes sprinkled with pyramids of varying size. All this is possibly useful for quantum device construction, but it is hard to control and is scarcely unexpected.

Now bury these pyramids under another layer of silicon, so that the tips lie under several atomic layers. Then start to deposit the Ge-Si alloy again. The first surprise is that the new crystallites grow directly over the tips of those buried below. The second is that this is an inexact description only insofar as the new array of quantum dots

looks a little more orderly — there is a correlation between pyramids in the first and second layers, but also a *lateral* correlation between those within the second layer. Repeat this sequence of burial and new pyramid growth, and the new array is even more correlated — the pyramids look reasonably well aligned into rows, and of more uniform size. For those who do not trust the eye, a Fourier analysis of microscopic images shows a progression from a blurred ring to symmetric, sharpening spots.

These experiments, not yet published, were done by researchers at the University of Wisconsin and at AT&T Bell Laboratories. They set Tersoff, a theorist, thinking about the origin of the self-organization. The correlations between crystallite positions in successive layers are easily understood: they are a consequence of the strain imposed on the atomic planes of the silicon overlayer by the embedded Ge-Si crystals, which is maximal over their tips and decays to all sides. This strain, transmitted to the uppermost silicon layer, creates a higher surface energy and thereby enhances the possibility of crystallite nucleation.

But what of the gradual ordering and size uniformity? Tersoff first developed a two-dimensional model of the process, although later confirming that it works for three dimensions too. His calculations show that, if two buried pyramids are close enough together, their strain maxima coalesce into a single peak. That is not astonishing, but it is enough almost by itself to generate increasing order in successive layers of crystallites. The only other factor needed is that between widely spaced buried pyramids there is a very shallow strain maximum more or less equidistant from the two, so a 'fresh' crystallite is most likely to nucleate halfway between them. In this way, large gaps in an underlying layer are filled in the succeeding one.

Tersoff shows that these two factors act as a kind of band-pass filter for crystallite nucleation, preventing two from forming

too close together (as the strain maxima from below coalesce) or staying too far apart (as new crystallites will fill in the gaps). Very quickly (after about six iterations), an initially random distribution of pyramids turns into a fairly regular array. But the distribution never becomes fully ordered, because the band-pass filter gives diminishing returns, always allowing some noise to remain in the pass band.

If perfectly ordered quantum dots are what you need, Mounji Bawendi and colleagues at the Massachusetts Institute of Technology have a better solution, as related by Chris Murray from Bawendi's group. He is to be commended for resisting the temptation to pass off as 'self-organization' a process that is in fact colloidal crystallization — but the degree of organization here is nevertheless phenomenal. Bawendi's group have synthesized nanometre-scale clusters of crystalline semiconducting cadmium selenide of essentially uniform size and spherical shape, whose size dispersion is less than an atomic layer (*Science* **270**, 1335–1338; 1995). They prepare the crystals under conditions in which nucleation and growth are closely controlled to give particle sizes within the range of 2 to 15 nanometres, and then passivate the surfaces by attaching organic substituents — trialkylphosphines, which give each cluster a hairy coat of hydrophobic alkyl tails. These 'soft' balls will coalesce from dispersions in polar alcohols, because of colloidal interactions, forming perfectly regular, hexagonally packed superlattices — colloidal crystals that are large and robust enough to be picked up with tweezers.

Colloidal crystallization is well known for polymer spheres of micrometre size, dispersions of which settle into iridescent crystals which are loosely bound and easily disrupted by shear. And aggregation of silica spheres in the natural environment leads to opal, with its pearly lustre. But the MIT group have done far more than simply shrink to the nanoscale conventional stuff from the microscale. Cadmium selenide clusters this small have their electronic band structure modified by quantum size effects — the well-known particle-in-a-box idea — bringing their bandgap into the visible part of the spectrum and making it continuously tunable as a function of size. So these materials are potentially electroluminescent media for light-emitting devices; indeed, CdSe clusters have been used in such (see V. L. Colvin *et al.* *Nature* **370**, 354; 1994). The control that Bawendi and colleagues have over size (and thus absorption and emission wavelength) and crystallization properties (the superlattice spacing can be varied by altering the alkyl chain length, for example) promises to make these nanocluster arrays a serious contender in the burgeoning field of display devices. □

Philip Ball is an associate editor of Nature.

*Fall Meeting of the Materials Research Society, Boston, USA, 27 November – 1 December 1995.

When colours move

Patrick Cavanagh

MOVING objects are typically distinct from their backgrounds in many respects, luminance, colour, texture and depth, for example. It would make sense for our visual systems to analyse each of these attributes to determine if an object has moved, but the evidence until now has favoured the view that only luminance counts and that colour in particular is ignored. The paper by Cropper and Derrington on page 72 of this issue¹ overturns this widely held belief and shows that colour is an independent contributor to motion perception.

It was more difficult than might have been expected to come up with convincing evidence of this independent role of colour, as the colour and luminance pathways in visual processing are not neatly separated but, rather, are richly interconnected. Overall, the similarities of motion responses to colour and to luminance stimuli now look more interesting than their differences. They point to the processing of several independent measurements by the brain as a strategy for improving the reliability of visual analyses.

Cropper and Derrington's experiment examined the nature of the initial, directionally selective neurons of the visual system. A vast array of these units, each monitoring a small local area, directly signals the local motions in the retinal image in the same way that other arrays of units decompose local orientation, colour and binocular disparity. Do these low-level detectors come in different types, one type for moving luminance borders and another for moving colour borders? And if so, why?

The luminance pathway (nominally composed of units which add red-sensitive and green-sensitive signals from the cone cells in the retina) has always been assumed to be the main site of motion-detecting units. However, any departure from the simplest linear combination of cone signals will create a response to colour stimuli in these luminance-based units. And, of course, the visual system is a happy home to many such departures. For example, when presented with an alternating pair of colours, units in the magnocellular pathway, the putative site of the luminance-based motion detectors, respond twice, once to each transition, and are not silenced at any ratio of relative luminance between the two colours^{2,3}. This 'cross-over' distortion results in a frequency-doubled response to colour gratings in these units. Moreover, the luminance pathway responds to red-sensitive and green-sensitive cone inputs with large and variable phase shifts, making nominally chromatic stimuli (red- and

green-sensitive responses 180° out of phase) strong contributors to a luminance response. Perhaps because of these factors, physiological studies of individual units of the primary visual cortex and the midtemporal cortex, the part of the brain mainly implicated in higher-order processing of motion, have revealed responses to coloured stimuli⁴⁻⁶. But none found directionally selective units that respond to colour and not luminance stimuli.

What led to the idea that colour did not contribute to motion? Early studies of motion perception by human observers showed that in some situations moving coloured stimuli, in particular equiluminous random-dot patterns, do not produce impressions of motion⁷. Simple coloured stimuli such as bars or gratings do produce impressions of motion but the pattern may appear to be significantly slowed or even stopped⁸. Basically, the threshold for seeing a colour pattern is lower than the threshold for seeing it move⁹⁻¹¹, so it is possible to present a drifting colour stimulus that is clearly seen but appears to be stationary. The motion seen for coloured stimuli was often claimed to be a result of a higher-level tracking system. This high-level system may have evolved separately to provide a parallel analysis of motion based on tracking the position of any visible feature, colour included¹². Others thought that the motion for coloured stimuli was a result of residual responses in the luminance pathway.

However, several recent articles have claimed a surprisingly robust response to the motion of colour stimuli^{10,11,13}. The threshold for discriminating the direction of motion for slowly moving, low-spatial-frequency gratings was found to be as much as four times lower for colour gratings than for luminance gratings. The notion of residual response in the luminance pathway cannot explain these findings, nor those of Hawken *et al.*¹⁴, where effects of stimulus contrast on speed judgements were qualitatively different for luminance and colour stimuli. But in all of these studies the stimuli were presented for at least 250 milliseconds and as much as one second duration, time enough to locate a coloured feature and track it either with attention, or with eye movements, thus revealing its motion. In studies where such tracking would be difficult, thresholds for the motion of colour stimuli were as much ten times higher¹⁰. At these high contrasts, it is difficult to rule out residual responses in the luminance pathway.

But that is exactly what Cropper and Derrington were able to do. First of all,

their colour stimulus, a sinusoidal grating of red and green bars, made only one jump of 1/4 of a cycle, a jump size which would be an ambiguous 1/2 cycle to the frequency-doubled response of luminance-based units. Second, they masked their stimulus with a stationary luminance grating that strongly affected moving luminance tests but had virtually no effect on the colour test, thus demonstrating the independence of the colour-specific motion response from luminance-based mechanisms. Finally, an extremely brief, 17-millisecond presentation guaranteed that no feature tracking was possible.

Why should the brain bother to extract motion signals for colour and, especially, why should it go to the trouble of doing so independently of a similar analysis of luminance-defined borders? One reason is that many luminance borders in a scene arise from shadows and highlights, and the play of light and shadow on an object can be quite independent of the object's motion. Colour borders and their motion are more reliable indicators of an object's contour and trajectory. A parallel analysis of motion for both luminance and colour should improve motion extraction in a statistical sense as well. Interestingly, this advantage could only be exploited if the separate analyses are subsequently combined in some way. Indeed, the motion of chromatic and luminance gratings drifting in opposite directions¹⁵ can cancel each other, showing that the motion responses to luminance and colour are integrated at a subsequent site. The key point of Cropper and Derrington's paper is that it shows that, as required for optimal integration, the chromatic input to this integrated site is derived independently of the luminance input. □

Patrick Cavanagh is in the Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, Massachusetts 02138, USA.

1. Cropper, S. J. & Derrington, A. M. *Nature* **379**, 72-74 (1996).
2. Schiller, P. H. & Colby, C. L. *Vision Res.* **23**, 1631-1641 (1983).
3. Lee, B. B., Martin, P. R. & Valberg, A. J. *Physiol., Lond.* **414**, 223-243 (1989).
4. Hubel, D. H. & Livingstone, M. S. *J. Neurosci.* **10**, 2223-2237 (1990).
5. Dobkins, K. R. & Albright, T. D. *J. Neurosci.* **14**, 4854-4870 (1994).
6. Gegenfurtner, K. R. *et al. Visual Neurosci.* **11**, 455-466 (1994).
7. Ramachandran, V. S. & Gregory, R. *Nature* **275**, 55-56 (1978).
8. Cavanagh, P., Tyler, C. W. & Favreau, O. E. *J. opt. Soc. Am. A1*, 893-899 (1984).
9. Lindsey, D. T. & Teller, D. Y. *Vision Res.* **30**, 1751-1761 (1990).
10. Derrington, A. M. & Henning, G. B. *Vision Res.* **33**, 799-811 (1993).
11. Stromeyer, C. F. III, Kronauer, R. E., Ryu, A., Chaparro, A. & Eskew, R. T. Jr *J. Physiol., Lond.* **485**, 221-243 (1995).
12. Cavanagh, P. *Science* **257**, 1563-1565 (1992).
13. Gorea, A., Papathomas, T. V. & Kovacs, I. *Vision Res.* **33**, 2515-2534 (1993).
14. Hawken, M. J., Gegenfurtner, K. R. & Tang, C. *Nature* **367**, 268-270 (1994).
15. Chichilnisky, E.-J., Heeger, D. & Wandell, B. A. *Vision Res.* **33**, 2113-2125 (1993).

VOLCANISM

Taupo's atypical arc

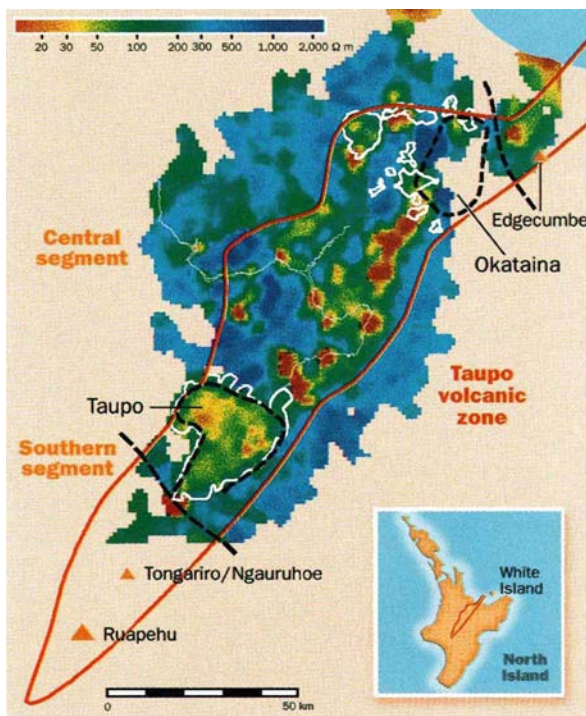
Colin J. N. Wilson

THE recent activity of Mount Ruapehu, at the southern end of the Taupo Volcanic Zone in New Zealand, has given locals their largest eruptions in 50 years, their first taste of volcanic smog ('vog' to aficionados), and a sharp reminder that theirs is a peculiarly active part of the planet. Although the Ruapehu eruptions may have been spectacular, they are dwarfed in size and violence by numerous prehistoric eruptions which have occurred elsewhere during the 2-million-year history of the Taupo Volcanic Zone (TVZ). This zone, the topic of a recent set of review papers¹, represents the southernmost, continental portion of a roughly 2,800-km-long volcanic arc associated with the subduction zone which stretches northwards through the Kermadec Islands to Tonga.

In subduction zones beneath continental crust, silica-poor basaltic magmas are generated in the wedge between the over-riding and downgoing plates. These mantle-derived magmas rarely reach the surface unmodified, but instead pond and stagnate in the crust, melt and incorporate crustal material and partially crystallize to generate lower-density, more silicic, more viscous intermediate magmas. Thus most arc volcanoes mainly erupt magmas of composition ranging from andesites to dacites, and do so frequently and in modest volume to build the archetypal large cone volcanoes like Ruapehu. In some arc volcanoes or arc segments, the magmatic processes that create intermediate magmas run to the extreme in producing rhyolite. Rhyolite magma is rich in silica, highly viscous and usually gas-rich, and is often erupted explosively to form widespread, sheet-like deposits, so that rhyolite volcanoes are not cones but unimpressive low-lying edifices often tens of kilometres across. Occasionally rhyolite is erupted in such large volumes (tens to thousands of cubic kilometres) that the ground surface above the drained subsurface magma chamber collapses to form a caldera.

The TVZ is divisible into three segments (see map). Two of these, to the north and south, are similar to many other continental arcs around the world, with andesite-dacite cone volcanoes and relatively low geothermal heat fluxes. In sharp contrast, in the middle of the TVZ is an

extraordinarily productive area of rhyolitic volcanism and huge geothermal heat fluxes which is one of the most unusual arc segments in the world. Other volcanic arcs release heat (carried by magma and geothermal fluids) at rates determined primarily by the speed of convergence in the subduction zone², and on this basis the



Contrast between where the geothermal heat is being released in the TVZ and the magmatic areas. The outline shows the TVZ with its three segments; the main young active rhyolite caldera volcanoes of Taupo and Okataina in the central TVZ; the main young active andesite-dacite cone volcanoes of the northern segment (White Island and Edgumbe); and the main young active andesite-dacite cone volcanoes of the southern segment (Tongariro/Ngauruhoe and Ruapehu). Superimposed colours represent the resistivity of country rocks (in Ωm) within about 300 m of the ground surface⁴. Areas of low resistivity (orange to red colours) indicate where hot mineralized geothermal water is rising to the surface or present at very shallow levels, and these areas are where the chief fluxes of geothermal heat occur in the central North Island (inset). Thermal fluxes outside the central area of TVZ are much lower. Note that the low resistivity zone near the Bay of Plenty coastline reflects intrusion of highly conducting salt water below the surface, not the presence of hot geothermal fluids.

'norm' for the TVZ would be about 1,200 MW thermal output for the 200 km onshore length of the zone. Within the 120×40 km central segment of the TVZ alone, the outputs of rhyolite magma ($0.28 \text{ m}^3 \text{ s}^{-1}$, equivalent to approximately 800 MW)³ and geothermal fluids (equivalent to $4,200 \pm 500 \text{ MW}$)⁴ mean the heat flux is more than 7 times the norm, a figure unrivalled by any other arc segment, and only equalled on Earth by major

'hotspot' areas like Yellowstone⁵, where rhyolite is produced at $0.13 \text{ m}^3 \text{ s}^{-1}$ accompanying a 5,300-MW thermal flux. The causes of the extraordinary thermal and rhyolite magma fluxes in the central TVZ are controversial, with no agreement except about the area's unusual, if not unique, nature.

The geothermal heat flux in the central TVZ can be mapped most easily by electrical resistivity techniques⁴, where hot, mineralized geothermal fluids produce anomalous areas of low resistivity near the

Earth's surface (see map). Most of these areas are related to active geothermal fields, used for power generation and tourism. The intensity of the geothermal flux and the measured temperature gradients are too high to reflect simply a more-efficient focusing of the conductive thermal gradients seen in conventional arcs, but whether the real cause of the excess flux is magma injection into the lower crust⁶, or plastic crustal deformation⁷, or both, is controversial. Nevertheless, there are other curiosities, such as the fact that many of the geothermal areas occur in places where volcanic activity has been dormant for 100–200,000 years; if magmatism is occurring at depth, why is active volcanism absent? Conversely, the modern rhyolitic caldera volcanoes of Taupo and Okataina (see map) have relatively low geothermal heat flows associated with them.

Although the central TVZ is the most productive and active rhyolitic system on the planet, the origin(s) of the rhyolite magmas have not been satisfactorily explained. As in other arcs, the most plausible ways of generating rhyolite magma in the TVZ are by having the magmatic crystallization processes run to their limit (thus raising the silica content of the residual magma) or by inducing crustal melting to produce rhyolite magma directly, or a combination of the two (termed AFC; assimilation then fractional crystallization). Assessing the relative roles of these or other more exotic processes is tricky, in large

part because we can only speculate on the nature and composition of the crust below the TVZ. Information from geothermal drillholes extends only to 2.8 km depth, and what lies between that level and the base of the crust at roughly 15 km is unknown. Several plausible candidate rock types have been put forward, but there is no good way of knowing which is most likely. Proponents of wholesale crustal fusion as a source for the rhyolites



Eruptive activity on Mount Ruapehu. Photography by I. A. Nairn, Institute of Geological and Nuclear Sciences.

have problems in reconciling all the geochemical characteristics of single rhyolite samples with those of any credible crustal source material. Proponents of crystallization models have equally knotty problems in explaining why their chosen process should have gone on to produce dominantly rhyolite, instead of andesite or dacite, and where the large volumes of crystalline material (complementary to the erupted rhyolite) are to be found. At present, combination (AFC) models are preferred^{6,7}, in which the mantle-derived basalts melt and incorporate a limited amount of crustal material (up to 25 per cent by mass), then the resulting andesitic magma evolves to rhyolite by crystallization. This combination approach neatly inherits and suffers from the same problems as both of the contributing mechanisms, and still fails to explain why rhyolite volcanism should be so geographically confined in an area so closely coincident with the area of high geothermal heat flux.

The central TVZ is unusual in other respects. The Yellowstone system (like many other rhyolitic systems in the western United States) shows cycles of activity that revolve around huge caldera-forming eruptions, on time scales of 10^5 – 10^6 years, which are inferred to tap long-lived magma chambers of enormous size⁸. In contrast, the TVZ has more frequent but moderate-sized caldera-forming events, on average every 40,000–50,000 years⁹. This higher eruption frequency is now being interpreted⁹ to reflect the thinned and intensively faulted nature of the crust below the central TVZ (which is rifting at rates of 7–18 mm yr⁻¹; ref.10), which does not permit the build-up of huge, long-lived magma chambers. For example, at Taupo volcano itself, rapid step-wise changes in composition of rhyolite eruptions over the past 20,000 years suggest that a large magma chamber may not even be present below the volcano for much of the time¹¹.

This leads to some intriguing new possibilities. First, if rifting is important in controlling rapid release of magma stored in the crust, does this mean that timings of eruptions at individual volcanoes in the TVZ may reflect tectonic, not magmatic, controls¹²? Second, if the high frequency of TVZ eruptions results in 'snapshots' of the magmatic processes at intervals of decades to a few thousand years, much shorter than is considered the norm⁸, does this permit us to disentangle the processes forming the rhyolite from those modifying the rhyolite as it is held in the crust before eruption?

The central TVZ has been known for over 130 years to be an exceptionally vigorous area of rhyolitic volcanism¹³. Although in the mean time an impressive body of observations has been accumulated on the TVZ, the search for the causes of its unique features look likely to exercise geologists for some years to come. □

Colin J. N. Wilson is at the Wairakei Research Centre of the Institute of Geological and Nuclear Sciences, Private Bag 2000, Taupo, New Zealand.

1. Simmons, S. F. & Weaver, S. D. (eds) *J. Volcanol. geotherm. Res.* **68**, 1–238 (1995).
2. Hochstein, M. P. *J. Volcanol. geotherm. Res.* **68**, 117–151 (1995).
3. Wilson, C. J. N. et al. *J. Volcanol. geotherm. Res.* **68**, 1–28 (1995).
4. Bibby, H. M., Caldwell, T. G., Davey, F. J. & Webb, T. H. *J. Volcanol. geotherm. Res.* **68**, 29–57 (1995).
5. Smith, R. B. & Braille, L. W. *Geol. Surv. Wyoming Mem.* **5**, 694–754 (1993).
6. McCulloch, M. T., Kyser, T. K., Woodhead, J. & Kinsley, L. *Contrib. Mineral. Petrol.* **115**, 303–312 (1994).
7. Graham, I. J., Cole, J. W., Briggs, R. M., Gamble, J. A. & Smith, I. E. M. *J. Volcanol. geotherm. Res.* **68**, 59–87 (1995).
8. Smith, R. L. *Geol. Soc. Am. Spec. Pap.* **180**, 5–27 (1979).
9. Houghton, B. F. et al. *Geology* **23**, 13–16 (1995).
10. Darby, D. J. & Williams, R. O. *N. Z. J. Geol. Geophys.* **34**, 127–136 (1991).
11. Sutton, A. N., Blake, S. & Wilson, C. J. N. *J. Volcanol. geotherm. Res.* **68**, 153–175 (1995).
12. Wilson, C. J. N. *Phil. Trans. R. Soc. Lond. A* **343**, 205–306 (1993).
13. von Hochstetter, F. *Geologie von Neu Seeland* (1864; Engl. transl. Fleming, C.A., Government Printer, Wellington, New Zealand, 1959).

DAEDALUS

Deep treasure

HEAVY metals play a crucial part in technology. Gold for circuit boards, platinum for catalysts and crucibles, tungsten for lamp filaments and uranium for nuclear power, all are rare and expensive. They are also all about twenty times as dense as water, seven times as dense as rock. And this, says Daedalus, is why they are rare. Geologically, our Earth is a viscous fluid; over its billion-year history these heavy metals have almost completely sunk down into its core.

The heavy metals do occur in sea water, though in tiny concentrations: about 1 part in 10^{11} for gold, for example. But gold and its compounds are much denser than sea water. They will sink in the oceans just as they do in the Earth. Their concentrations should rise exponentially with depth, like a solution of high-molecular-weight protein in an ultracentrifuge. Daedalus estimates that 10 km down, there should be at least a thousand times as much gold as at the surface, and maybe many millions of times, depending on the molecular weight of the dissolved gold compound. Abyssal water may well be richer than the gram per tonne of typical goldmine ore.

So DREADCO oceanographers are fitting a drilling barge with ten kilometres of polypropylene pipe, to reach down into the deepest trenches of the ocean and pump up the water. The pipe will be weighted to exactly neutral buoyancy so that it will neither load the vessel down nor be stressed by its own weight. As the water comes aboard, it will be chemically processed to extract the heavy metals: not only gold, but platinum, palladium, rhodium and many others. The operation should be hugely profitable.

Mercury, the only liquid heavy metal, may be a special case. It often occurs 'native' as the pure liquid. Over geological time this too should have found its own level and sunk to the bottom of the ocean deeps. Huge pools of mercury must lurk under the ooze of the deep-sea trenches, waiting to be pumped up. Furthermore, they should have amalgamated and absorbed a rich haul of the other bottom-dwelling heavy metals. Sadly, no pump could suck up a ten-kilometre column of mercury. But Daedalus recalls that grinding mercury with a fat can divide it stably into tiny fat-coated globules. Mercurial ointments for syphilis were made this way, as was the famous, and famously hazardous, 'Blue Pill' purgative. The DREADCO team are designing a deep-sea ultrasonic agitator fed with lard, to break the mercury into an emulsion of such globules. The upward rush of pumped water should then entrain them and waft them effortlessly to the surface.

David Jones

Analgesic effects of myrrh

SIR—Myrrh is a natural compound secreted by shrubs of the genus *Commiphora* of the Burseraceae. It is common in northeast tropical Africa, and is composed of essential oils, water-soluble gums and alcohol-soluble resins¹. In antiquity, myrrh was used by the Egyptians for embalming and by the Jews as anointing oil. Hippocrates recommended myrrh for sores, and the Romans used it for treating mouth and eye infections, coughs and worm infestations^{2,3}. In St Mark's Gospel, "vinum murratum", wine with myrrh, was offered to Christ before crucifixion. In later centuries, myrrh was an essential component of the pharmacopoeia.

To investigate its medicinal properties, we administered a suspension of ground commercial myrrh by gavage to mice and measured the latency of their pain reaction (paw licking) when placed on a 52 °C metal plate. Mice given a 10% suspension of ground myrrh in saline (10 ml per kg) 15 min after the administration, had a licking latency time of 19.4±2.4 s, compared with controls administered saline (14.4±0.6 s, $P<0.01$). We then identified the constituents of myrrh with analgesic activity from *Commiphora molmol*. A hexane extract containing the analgesic activity was separated in different fractions by silica gel column chromatography, followed by semi-preparative high-pressure liquid chromatography.

Using nuclear magnetic resonance and mass spectrometry, we identified three sesquiterpenes (see *a* in figure). The most abundant compound (>90%) was furanoeudesma-1,3-diene; the remaining compounds were curzarene and furanodiene. These sesquiterpenes had been previously identified^{4,5} but their biological effects had not been described.

We injected the purified compounds intracerebroventricularly in mice at a dose of 1.25 mg per kg. Furanoeudesma-1,3-diene, given 30 min after administration, increased licking latency from the baseline value of 15±0.7 to 20.1±1.6 s ($P<0.01$), and curzarene increased it from 15.5±2.2 to 21±1.8 s ($P<0.01$), whereas furanodiene was ineffective (from 14.5±0.6 to 12.6±0.7 s).

For the more abundant furanoeudesma-1,3-diene, a dose of 50 mg per kg orally (p.o.) considerably reduced the number of writhes (abdominal muscle contractions) in mice after 0.6% acetic acid intraperitoneal (i.p.) administration. This effect was completely reversed by naloxone (*b* in the figure). Morphine (5 mg per kg, p.o.) had similar activity. Significant analgesia was also seen with the hot plate test after furanoeudesma-1,3-diene administration p.o. at the same dose (*c* in the figure).

We then tried to characterize the binding of furanoeudesma-1,3-diene to opioid receptors in brain membranes using a

nonspecific radioactive ligand, [³H]diprenorphine, and observed a dose-related [³H]diprenorphine displacement, although with a high inhibition constant (pK_i , 5.7±0.8 M; ±s.d.).

In conclusion, two sesquiterpenes present in myrrh have analgesic effects blocked by naloxone, indicating an interaction with brain opioid mechanisms. This could explain the use of myrrh as a pain killer in ancient times. Its use for analgesia may later have been dropped and replaced by opium derivatives, given the presence in myrrh of other compounds with unknown or unfavourable pharmacological activity⁶. Furanoeudesma-1,3-diene could still have some medicinal applications, although its action on the central opioid pathways would limit its practical usefulness.

**Piero Dolara, Cristina Luceri
Carla Ghelardini, Claudia Monserrat
Silvia Aiolfi, Francesca Luceri
Maura Lodovici**

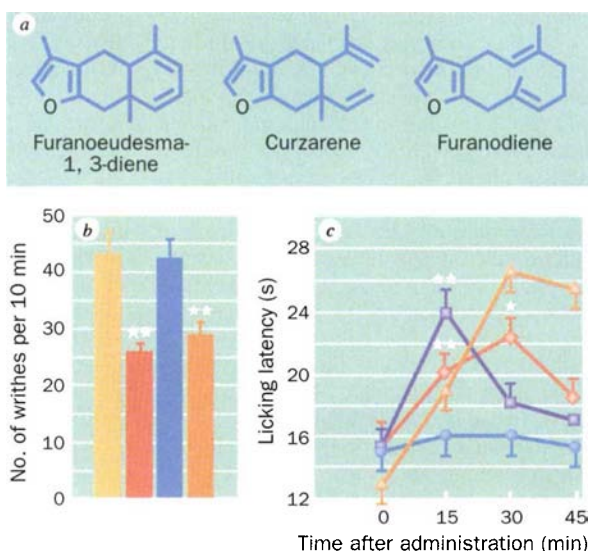
Department of Pharmacology,
Stefano Menichetti

Department of Organic Chemistry,
Maria Novella Romanelli

Department of Pharmaceutical Sciences,
Faculty of Pharmacy,
University of Florence,
50134 Florence, Italy

1. Tucker, A. O. *Econ. Bot.* **40**, 425–433 (1986).
2. Hilson, R. M. J. *R. Soc. Med.* **81**, 541–543 (1988).
3. Michie C. A. & Cooper, E. J. *R. Soc. Med.* **84**, 602–605 (1988).
4. Brieskorn, C. H. & Noble P. J. *Med. Pl. Res.* **44**, 87–90 (1983).
5. Brieskorn, C. H. & Noble P. *Phytochemistry* **22**, 187–189 (1983).
6. Claeson, P. et al. *Planta Med.* **57**, 352–356 (1990).

a, Structures of sesquiterpenes from myrrh tested for analgesic activity. *b,c*, Analgesic effect of furanoeudesma-1,3-diene and morphine in the hot plate and writhing test. The substances were administered by gavage or i.p. in a volume of 10 ml per kg of corn oil (for furanoeudesma-1,3-diene) or saline (for morphine and naloxone). Male Swiss-albino mice weighing 20–25 g were used. *b*, Writhing test: 25 min after test compound administration, 0.6% acetic acid (10 ml per kg) was injected i.p.; after an additional 5 min the number of abdominal muscle contractions (writhes) was scored for 10 min. Naloxone was administered 15 min after the test compounds. Yellow bar, corn oil (10 ml per kg, p.o.); red, furanoeudesma-1,3-diene (50 mg per kg, p.o.); blue, furanoeudesma-1,3-diene (50 mg per kg, p.o.) and naloxone (1 mg per kg, i.p.); orange, morphine (5 mg per kg, p.o.). *c*, Hot plate assay: furanoeudesma-1,3-diene was administered by gavage and morphine was administered subcutaneously. Naloxone was administered i.p. immediately after gavage with furanoeudesma-1,3-diene. At 15, 30 and 45 min after the administration of the test compounds, mice were placed on an open aluminium chamber (diameter, 20 cm; height, 30 cm) maintained at 52 °C, observed for initial pain reactions (paw licking) and immediately pulled out of the chamber after the first pain reaction, the latency being recorded in seconds. Red curve, furanoeudesma-1,3-diene (50 mg per kg, p.o.); purple, furanoeudesma-1,3-diene (100 mg per kg, p.o.); blue, furanoeudesma-1,3-diene (100 mg per kg, p.o.) and naloxone (1 mg per kg, i.p.); orange, morphine (5 mg per kg, s.c.). Data are means ± s.e. ($n=10$). ** $P<0.01$; * $P<0.05$, by Student's two-tailed *t*-test.



Tracking bees with harmonic radar

SIR—Much of our knowledge of the high-altitude flight behaviour of insects has been derived from the use of pulse radars¹, and there are many instances where equivalent information about low-altitude flight would also be of considerable entomological value, host-finding behaviour by tsetse flies, foraging by bees and flight to pheromone sources by Lepidoptera to name but a few. Unfortunately, radar reflections from ground features (clutter) prevent this technique from being used to observe low-level flight, except where this occurs over extremely flat and bare terrain². We have therefore used the harmonic radar principle^{3,4} to develop a method of measuring the trajectories of low-flying insects over distances of hundreds of metres. We report here its first trial application, the observation of foraging flights by bumble bees and honey bees.

The technique requires that the target insect be 'tagged' with an electrically non-

FIG. 1 Harmonic-generating tag mounted on a bumble bee (*Bombus terrestris*). The tag consists of a low-barrier Schottky detector diode mounted at the centre of a 16-mm dipole antenna, and in parallel with a 3-nH inductor. The complete assembly weighed ~3 mg, that is, 1.5% of the bee's weight. Free-fall experiments on the tag alone indicated that at a fall speed of 5 m s^{-1} its (broadside) aerodynamic drag was ~0.8 mg. By contrast, the body drag of a bumble bee flying at 5 m s^{-1} is 20–30 mg (ref. 5), so the tag seems unlikely to add substantially to the overall drag that the bee has to overcome. The upwards-turning moment that it induces may, however, perturb the bee's flight pitch angle, and this could perhaps result in a significant increase in drag.

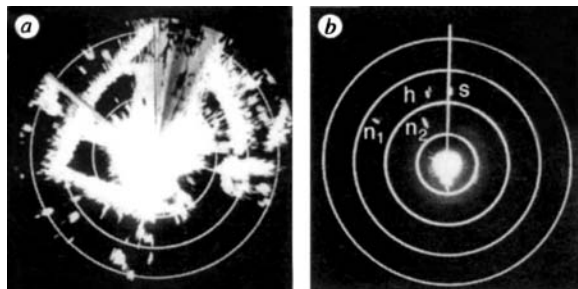
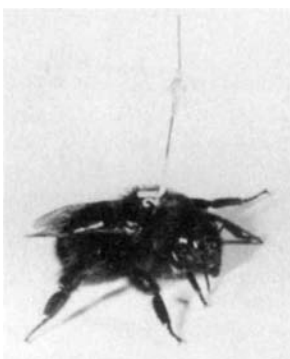


FIG. 2 a, The conventional radar mode display shows the 'clutter' from avenues of trees and ground features at our experimental site. b, The harmonic mode is activated, all ground clutter is eliminated, and the only fixed targets to appear are tags marking the two nests (n_1 and n_2) and the forage site (s). A honey bee is seen in flight (h), returning to n_2 . The distance from the centre to the outermost ring represents 463 m, and the direction of north is indicated by the vertical line; radar resolution is approximately $\pm 7 \text{ m}$ in range and $\pm 3.5 \text{ m}$ in azimuth.

linear device designed to re-radiate an harmonic of the radar signal which can be detected against even strong ground clutter. The energy to operate the tag is delivered by the illuminating radar, so no 'on-board' battery is required and extreme miniaturization is therefore possible (Fig. 1). Our radar transmits 25-kW pulses of 0.1 μs duration and 3.2 cm wavelength from a 1.5-m diameter paraboloid, and receives 1.6-cm-wavelength harmonic returns in a second 0.7-m-diameter dish mounted above the first. Both paraboloids rotate in azimuth at 20 revolutions per minute. The transmitting and receiving beams are of equal angular width and are approximately co-linear; they provided coverage over an altitude range of about 3 m and a range of 700 m. In our experiments, the altitude coverage was normally set to begin a few centimetres above ground level.

Three bumble bee colonies (*Bombus* spp.) and a small hive of honey bees (*Apis mellifera*) were placed 70 m and 250 m WSW of a plot of *Phacelia tanacetifolia*, which provided the major attractive forage source within close range of the colonies. After a few days, electronic tags were glued to some regular foragers from each nest. Tagged bees were released from either one of the nest sites, or from the

Phacelia plot. On release, they appeared as moving targets on the radar display (Fig. 2), and although they occasionally disappeared from the display when they landed or passed behind trees, round trips (nest to forage to nest) and one-way trips (from forage to nest) were satisfactorily tracked over distances of 50–250 m. Wind speed and direction were measured by a recording anemometer fixed at an altitude of 2 m, about 20 m north of position n_2 in Fig. 2. Subtraction of the wind vector from the bees' displacement vectors, measured by the radar, gave estimates of their air speeds. The mean value for air speed for all our tracks ($n = 41$) was $5.2 \pm 0.3 \text{ m s}^{-1}$, which is very similar to the values of 4–5 m s^{-1} found for bumble bee flights in long greenhouses (T. Wolf, personal communication).

Evidence that tagged bees could forage was provided by one honey bee which made a round trip, beyond our experimental area, and returned with orange pollen loads (probably from *Tripleurospermum inodorum*), and by the direct observation that tagged honey bees and bumble bees foraged on *Phacelia*.

Further experiments are needed to establish whether the tags significantly modify bee behaviour, but it seems likely that the harmonic radar technique could be used to study many aspects of bee flight and ecology. It also has potential for examining the patrolling routes of male bumble bees, and the searching flights of queens for over-wintering and nest sites. Application of the technique to many other insect species seems feasible, especially if we succeed in our attempts to miniaturize the tag still further.

J. R. Riley, A. D. Smith

D. R. Reynolds, A. S. Edwards

Natural Resources Institute Radar Unit,
North Site, Malvern,
Worcestershire WR14 1LL, UK

J. L. Osborne, I. H. Williams

N. L. Carreck, G. M. Poppy
IACR-Rothamsted, Harpenden,
Hertfordshire AL5 2JQ, UK

- Riley, J. R. A. *Rev. Ent.* **34**, 247–271 (1989).
- Loper, G. M., Wolf, W. W. & Taylor, O. R. *J. chem. Ecol.* **19**, 1929–1938 (1993).
- Shefer, J. et al. *Wireless World May* 117–122, June 199–202 (1974).
- Mascanzoni, D. & Wallin, H. *Ecol. Ent.* **11**, 387–390 (1986).
- Dudley, R. & Ellington, C. P. *J. exp. Biol.* **148**, 53–88 (1990).

Birth control for grey seals

SIR — Changing attitudes toward large marine and terrestrial mammals have fostered the need to develop new techniques for managing wild populations^{1,2}. In particular, increased pinniped abundance in many parts of the world has led to renewed pressure to control their population growth³. During the 1980s, advances in our understanding of zona pellucida glycoproteins led to their use in immunocontraception of females to limit the productivity of mammalian populations⁴. The practicality of this approach for use in wild populations was limited by the need for multiple administrations. Here we report the development of a single-administration immunocontraceptive vaccine for grey seals (*Halichoerus grypus*).

In January 1992, 207 adult female grey seals (132 14-yr-olds, 35 20-yr-olds and 40 21-yr-olds) on Sable Island, Nova Scotia (43° 55' N, 60° 00' W), were randomly allocated to either a control group (given a placebo vaccine) or an immunized group (see table). Pregnant grey seal females show a high degree of site fidelity to a breeding colony, facilitating recapture to determine birth rates and monitor antibody titres⁵. We are confident that most returning females in our study were located during daily surveys of the colony throughout the breeding season.

One year post-immunization, fewer immunized females (49%) gave birth than those that were given the placebo (67%), although this difference was not significant ($P = 0.1$; see table). Significantly fewer immunized females returned to give birth than did females given the placebo in year 2, when the full effect of the vaccine should be realized (10.8% as opposed to 70.5%; $P < 0.001$). Similarly, significantly fewer immunized females gave birth compared with the placebo group three years post-immunization ($P < 0.001$). The fertility of the immunized group was reduced by about 90% in both years 2 and 3.

No anti-soluble intact zona pellucida (SIZP) antibodies could be detected in

NO. OF INJECTED FEMALE GREY SEALS
RECAPTURED POST-IMMUNIZATION

Vaccine	Post-immunization (yr)			
	0	1	2	3
Placebo	105	70	74	66
Immuno-contraceptive	102	50	11	9
χ^2	—	2.7	45	41
P	—	0.1	<0.001	<0.001

Immunized females were given a single injection of SIZP (100 μg) encapsulated in multilamellar liposomes (M. M., US Patent 4,485,054) suspended in saline (0.5 ml) and emulsified in Freund's complete adjuvant (0.5 ml). Controls received a placebo vaccine lacking SIZP.

sera from the placebo group, whereas one year post-immunization, the average titre of the 50 recaptured seals immunized with the contraceptive vaccine was $25 \pm 26\%$ of the reference serum. Contrary to expectation, 7 of 8 females with two or more recaptures had stable or increasing anti-SIZP titres during the three-year monitoring period. This suggests the immunocontraceptive vaccine could be effective for more than three years.

**R. G. Brown, W. C. Kimmins
M. Mezei, J. Parsons
B. Pohajdak**

*Department of Biology,
Dalhousie University, Halifax,
Nova Scotia B3H 4J1, Canada*

W. D. Bowen

*Department of Fisheries and Oceans,
Marine Fish Division,
Bedford Institute of Oceanography,
Dartmouth, Nova Scotia B2Y 4A2, Canada*

1. Bonner, W. N. *Biol. J. Linn. Soc. Lond.* **38**, 53–60 (1989).
2. Garrott, R. A. *Wildl. Soc. Bull.* **19**, 52–58 (1991).
3. Butterworth, D. S. *S. Afr. J. Sci.* **88**, 414–416 (1992).
4. Service, R. F. *Science* **266**, 1480–1481 (1994).
5. Pomeroy, P. P. et al. *J. Zool.* **233**, 429–447 (1994).

Illusory motion from shadows

SIR — At least since the time of da Vinci, it has been known that shadows can give a qualitative sense of depth^{1,2}. Moving shadows, however, have been ignored as a source of information about the motion of three-dimensional objects in both biological and computational theories of vision. Rather, shadows have been considered a source of ambiguity. We show here that shadows can have a strong influence on the perceived three-dimensional motion of objects. The movement of an object's shadow induces apparent motion in depth of the object, even when the image of the object itself is stationary.

We used a computer graphics animation consisting of a checkerboard with a

central square in front of it, and a shadow moving towards and away from the square (Fig. 1). The size of the square was constant over time, thereby setting the three-dimensional motion cues of object expansion and contraction to indicate no motion in depth. The lack of any translatory motion of the square in the image provided further evidence, based on the assumption of a general viewing position, that the square was stationary in three dimensions. Despite the presence of these strong cues, all observers of the animation reported a strong perceptual impression of a square moving in depth.

If the apparent motion is specific to the processing of cast shadows, reducing the likelihood of labelling a region as a shadow should reduce the probability of obtaining the effect. We tested this prediction by measuring how often observers reported seeing the illusion under four different illumination conditions, generated by crossing two pairs of illumination properties: extended or point light source (shadows with penumbrae or sharp edges), and the light source from above or from below (Fig. 2).

Perceived motion was more likely when the shadow had a penumbra or when illumination was from above. The optimal shadow condition (fuzzy shadow below the square) replicated our initial observations, with all 15 observers reporting a perceived motion in depth on initial viewing. The worst shadow condition (sharp shadow above the square) showed the weakest motion induction effect, with only 7 out of 15 observers reporting the effect on initial viewing.

The kind of brain computation suggested by our findings resembles a process in which the visual system seeks a rational interpretation of the image sequence based on knowledge of how images could be formed from objects, their spatial relations, the illumination and viewpoint, together with prior assumptions about the nature of the world³. Theoretically, the visual interpretation is ambiguous. The

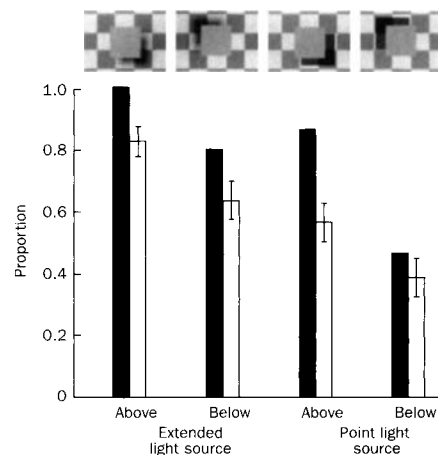


FIG. 2 The proportion of times observers reported the central square patch to be moving in depth for extended and point light sources from above or below. The black bars show the proportion (out of 15) of motion-in-depth judgements reported on first observations only. There is a significant advantage of an extended light over a point source ($z = 2.28$, $P < 0.02$), and of light from above as opposed to below ($z = 3.028$, $P < 0.002$). The open bars show the proportion of motion-in-depth judgements averaged over repeated observations (240 observations per condition). The error bars show 95% confidence intervals. Sixty observers were tested four times for each illumination condition, giving 16 responses per observer.

kinematics of the shadow could be due to a light source moving, or to motion of the square itself. The strong apparent motion of the square in depth suggests that the visual system resolves this ambiguity by assuming that light sources are usually stationary. Previous studies have shown that the perception of shape assumes that illumination is from above⁴, and that shadows are dark⁵. Our results show that depth perception takes advantage of the fact that light sources are generally stationary and above, thereby obviating the need for the computational resources required to estimate light source position and motion precisely.

Daniel Kersten

*Department of Psychology,
University of Minnesota,
Minneapolis, Minnesota 55455, USA*

David C. Knill

*Department of Psychology,
University of Pennsylvania,
Philadelphia, Pennsylvania 19104, USA*

Pascal Mamassian

*Department of Psychology,
New York University,
New York, New York 10003, USA*

Isabelle Bühlhoff

*Max Planck Institute for Biological
Cybernetics,
D-72076 Tübingen, Germany*

1. da Vinci, L. *Notebooks of Leonardo da Vinci* (Dover, New York, 1970).
2. Yonas, A., Goldsmith, L. T. & Hallstrom, J. L. *Perception* **7**, 333–341 (1978).
3. Gregory, R. L. *The Intelligent Eye* (McGraw-Hill, New York, 1970).
4. Ramachandran, V. S. *Nature* **331**, 163–166 (1988).
5. Cavanagh, P. & Leclerc, Y. G. *J. exp. Psychol. hum. Percept. Performance* **15**, 3–27 (1989).

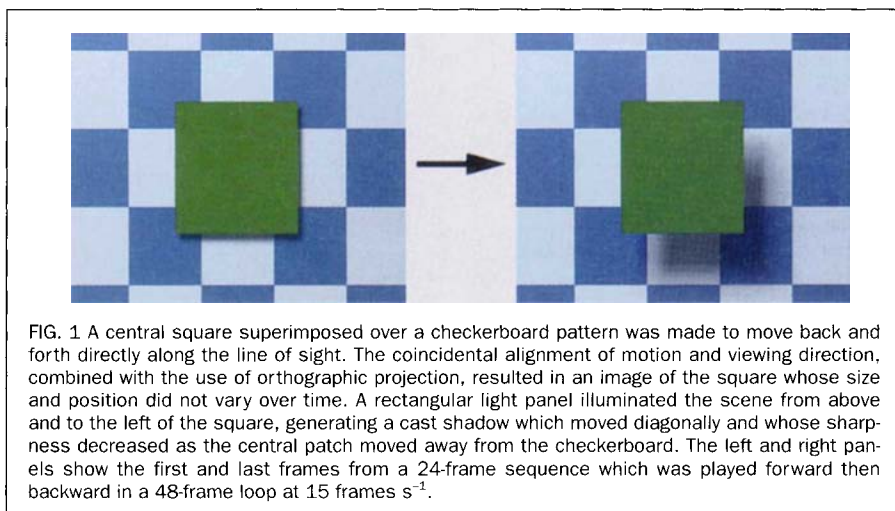


FIG. 1 A central square superimposed over a checkerboard pattern was made to move back and forth directly along the line of sight. The coincidental alignment of motion and viewing direction, combined with the use of orthographic projection, resulted in an image of the square whose size and position did not vary over time. A rectangular light panel illuminated the scene from above and to the left of the square, generating a cast shadow which moved diagonally and whose sharpness decreased as the central patch moved away from the checkerboard. The left and right panels show the first and last frames from a 24-frame sequence which was played forward then backward in a 48-frame loop at 15 frames s^{-1} .

Fossilized theropod soft tissue

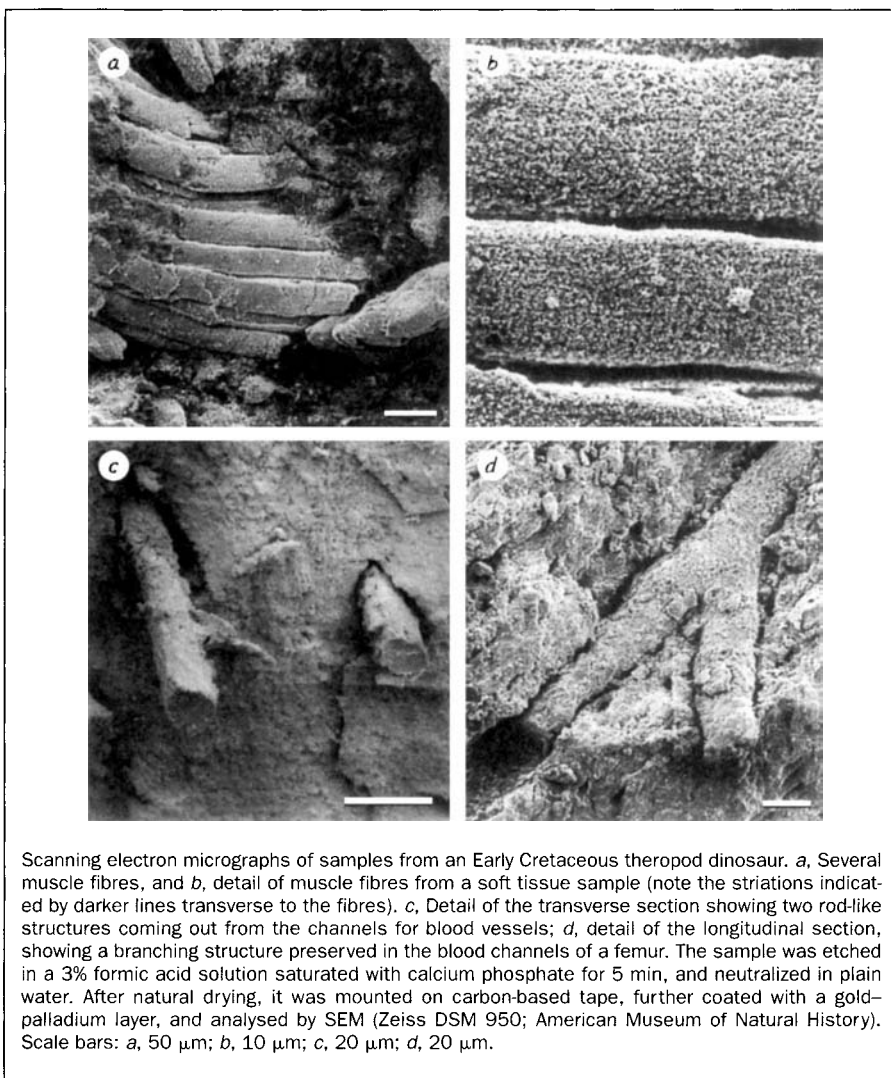
SIR—I report here the discovery of a small non-avian theropod dinosaur with extremely well-preserved soft tissue, including muscle fibres, in Cretaceous rocks of Brazil. Such well-preserved soft tissue calls attention to the anatomical details that can be preserved in fossils of terrestrial vertebrates.

The specimen (MCT 1502-R; Departamento Nacional da Producao Mineral, Rio de Janeiro) is preserved in a calcareous concretion, typical of those from the Santana Formation (Romualdo Member, Albian) of the Araripe Basin, north-eastern Brazil¹. This area is rich in fossils, but remains of dinosaurs are rare². The material comprises the pelvis, hindlimbs and some caudal vertebrae. The presence of three comparatively large pedal digits, of which the third is the largest, supports its placement in the Theropoda³. Preliminary comparisons suggest maniraptoran affinities.

Soft tissue is preserved in several parts of the nodule. It can be distinguished from the matrix and bones by the bright colours and different texture, except near the unguals, where the replaced horny covering is darker than the bone. The epidermis is very thin (~0.04 mm), and formed mostly by irregular quadrangles bordered by deep grooves, presenting a criss-cross pattern. No evidence was found of any structure covering the skin, such as dermal ossicles, scales or feathers, which should be preserved if they were originally present.

A sample of the soft tissue in the femur-tibia-fibula area was removed and analysed by scanning electron microscope (SEM). On the dorsal view, fibrous structures could be identified. They are polygonal in transverse section and in the area examined their diameters range from 30 to 50 µm. A carbon-coated section was submitted to energy dispersive spectroscopy for elemental characterization, which showed that those structures are preserved as calcium phosphate. They represent striated muscle fibres, some of which even show partial striations (*a* and *b* in the figure). Phosphatized soft tissue from the Santana Formation were previously reported in fossil fishes⁴ and pterosaurs⁵⁻⁷. It has also been replicated in laboratory experiments⁸.

Longitudinal and cross-sections of a femur fragment were also examined by SEM. The transverse section of this bone shows channels for blood vessels (diameter 20–25 µm) and lacunae for osteocytes (diameter ~5 µm). Rod-like structures have been preserved in some channels (*c* in the figure). Some show a



Scanning electron micrographs of samples from an Early Cretaceous theropod dinosaur. *a*, Several muscle fibres, and *b*, detail of muscle fibres from a soft tissue sample (note the striations indicated by darker lines transverse to the fibres). *c*, Detail of the transverse section showing two rod-like structures coming out from the channels for blood vessels; *d*, detail of the longitudinal section, showing a branching structure preserved in the blood channels of a femur. The sample was etched in a 3% formic acid solution saturated with calcium phosphate for 5 min, and neutralized in plain water. After natural drying, it was mounted on carbon-based tape, further coated with a gold-palladium layer, and analysed by SEM (Zeiss DSM 950; American Museum of Natural History). Scale bars: *a*, 50 µm; *b*, 10 µm; *c*, 20 µm; *d*, 20 µm.

different texture from the matrix and bone: the inner portion is smooth and covered by a rougher outer layer. Longitudinal sections show these structures branching off to nearby channels (*d* in the figure). The outer layer of those structures is rich in calcium phosphate, although not as rich as the bone. Both differ from the calcium carbonate matrix.

These structures can be interpreted as mineralizations filling the channels for capillary blood vessels of the bone, or they might represent the replacement of the blood vessels themselves. The fact that some of these structures have a smaller diameter than the channels in which they are preserved prevents the exclusion of the second hypothesis.

Most dinosaur soft tissue reported is preserved as impressions (for example, refs 9–11), and in one case as a kerogenous film of the skin¹². None approaches the condition of the material examined here, where not only the skin but also muscle fibres and other soft tissues are preserved in three dimensions. Careful collecting and preparation of other dinosaur specimens can provide more such material and might contribute to a

better understanding of the biology of these extinct animals.

Alexander W. A. Kellner

Fellow CNPq, Brasilia, and Department of Vertebrate Paleontology, American Museum of Natural History, New York 10024-5192, USA

1. Maisey, J. G. (ed.) *Santana Fossils, An illustrated Atlas* (TFH, New Jersey, 1991).
2. Campos, D. A. & Kellner, A. W. A. in *Santana Fossils, An Illustrated Atlas 372–375* (TFH, New Jersey, 1991).
3. Gauthier, J. *Mem. Calif. Acad. Sci.* **8**, 1–55 (1986).
4. Martill, D. M. *Palaeontology* **31**, 1–18 (1988).
5. Martill, D. M. & Unwin, D. M. *Nature* **340**, 138–140 (1989).
6. Kellner, A. W. A. *Annuar. Inst. Geoci.* **1989**, 86–106 (1990).
7. Kellner, A. W. A. *Acta geol. Leopold.* **39**, 615–625 (1994).
8. Briggs, D. E. G. *et al. J. geol. Soc. Lond.* **150**, 1035–1038 (1993).
9. Osborn, H. F. *Mem. Am. Mus. nat. Hist.* **1**, 31–54 (1912).
10. Bonaparte, J. F. & Powell, J. E. *Mem. Soc. géol. Fr.* **139**, 19–28 (1980).
11. Bonaparte, J. F. *et al. Contr. Sci.* **416**, 1–42 (1990).
12. Martill, D. M. *Mod. Geol.* **16**, 61–68 (1991).

Charcot's or ALS

Cleveland *et al.* wish it to be known that all reference to amyotrophic lateral sclerosis (ALS) as Charcot's syndrome in the title and text of their recent Scientific Correspondence¹ was inserted by the editors of *Nature*.

1. Cleveland, D. W. *et al. Nature* **378**, 342–343 (1995).

Emergent macrosimplicity

Ian Stewart

At Home in the Universe: The Search for Laws of Self-Organization and Complexity.
By Stuart Kauffman. Oxford University Press/Viking: 1995. Pp. 321. \$25, £20.

THIS is a courageous book. It has a distinct spiritual focus, not in the religious sense, but in the sense of 'why are we here?' When coupled with an unorthodox message that is unavoidably critical of much mainstream science, the combination is dynamite; and that is what makes it a courageous book.

The topic is complexity theory *à la* Santa Fe Institute, with an emphasis on developmental and evolutionary biology. The conventional biologist sees evolution as the result of random mutations in DNA sequences, mediated by natural selection, and abhors suggestions that evolution has any kind of inherent direction. Similarly, the conventional biologist sees development as a program written in DNA, a simple matter of assembling the right proteins. Stuart Kauffman argues that, in both areas, the conventional biologists have got the wrong end of the stick. He tells us that in adaptive complex systems a great deal of order and structure is often obtained 'for free'. It isn't built in; it emerges. He says that such systems can self-organize and self-complicate in a way that makes no sense to much conventional theorizing. He looks for a new kind of mathematics, one capable of dealing with emergence, that fits the breathtaking diversity of the natural world far better than the sterile rundown predicted by thermodynamics. And he thinks he knows where such a theory might be found.

Lost perspectives

To my mind, the conventional view of evolution loses an important perspective because it places too much emphasis on the least interesting features. Imagine a theory of the flow of water founded on the molecular nature of the liquid rather than the conventional continuum model favoured by fluid dynamicists. Suppose further that the only observations available are those of water in its natural habitat — streams and rivers, rainstorms and lakes — and that the surrounding landscape is invisible. We observe the water and try to explain its patterns as it meanders past a randomly strewn boulder or drops off an eroded ledge in a waterfall. And the story we tell is one of contingency and selection. At root, the flow of water is merely the random wanderings of molecules. There is no purpose, no goal, no order; just stochastic jiggling. But as the molecules jiggle around, some of them

tend to accumulate, whereas others tend not to. So the first kind do accumulate, whereas the second rapidly jiggle themselves elsewhere until eventually they accumulate too. As a consequence of this random jiggling plus selection based on the purely contingent factor of accumulation, the mass of water moves. It could, it seems, go anywhere.

But things look very different when we can see the landscape. The places where water molecules tend to accumulate are contingent on their surroundings, for sure, but once we know the surroundings we can predict what will happen. Water accumulates where the potential energy is lower — in short, it runs downhill. The geography of the surrounding landscape imparts a definite direction. The random jiggings are important for only one reason: they render the water fluid, so that it can flow into the lower regions. Without the random jiggings, motion would be only potential, but with them it becomes actual. And although the mechanism whereby individual molecules select where to accumulate is also contingent, it inherits direction from the landscape. (Individual molecules can climb uphill, but on the whole they don't.)

Evolutionary biologists seem to have the wrong view of physics. They think that the flow of water is deterministic, and that the partial differential equations used by fluid dynamicists are actually implemented by the fluid itself. They thus derive a wholly misleading contrast between Darwinian principles and physical laws, and this confuses them when they start thinking about the global constraints that affect evolution. 'Goal' and 'purpose' are not the right words, to be sure, but there are dynamic effects in evolution, resulting from constraints. Mutations make phenotypes fluid enough to change, selection implements particular changes preferentially, but the overall result is more like water flowing through a landscape. It is an invisible landscape, however, formed out of the nearby 'potential' phenotypes and constrained by context — a mathematical 'phase space'. It includes not only what happens, but what could have happened instead.

It sounds metaphysical, but such imagery lies at the core of how physicists and mathematicians currently think about all dynamics. To them, phase space is just as real as ordinary space — it makes itself

felt by constraining the potential dynamics into the behaviour that we actually observe.

What makes evolution particularly tricky is that the phase space is not fixed. It itself evolves, in response to the organisms that live in it. In *The Collapse of Chaos*, Jack Cohen and I coined the term 'complicity' for this kind of coevolution of content and context, but the words and concepts have not really been pinned down yet. We will never find them unless we recognize that there is some kind of global dynamic to evolution, and start trying to explore it. And that's one of the things that complexity theory is about: it shows that systems with a highly complex microstructure typically develop recognizable macrodynamics — usually in an emergent manner.

Topological constraints

One of Kauffman's basic examples is the Boolean network, a circuit of little light bulbs that are either on or off and that affect each other in an ever-changing way. Imagine a network with 100,000 bulbs. Mathematically such a system has only finitely many states — but vast, $2^{100,000}$ or about $10^{30,000}$ — so the only possible long-term dynamics is a periodic cycle. The question is: how long is the cycle? Short cycles are structured, very long ones might as well be random. Now, if all the bulbs connect together then the typical cycle has length about $10^{15,000}$. But if each connects to only two, then the typical cycle contains only 317 states. "I hope this blows your socks off", says Kauffman. Mine are well on their way to the Sahara. What this simple example tells us is that topology constrains dynamics. It quantifies one case where vast microcomplexity gives rise to emergent macrosimplicity.

Kauffman's main message is that most of biology is like this. The path of evolution is far less contingent than Stephen Jay Gould in *Wonderful Life* would have us believe, because it is constrained by the topology of its own phase space. The same is true of development: the growing organism builds its own phase space along with its cellular structure.

Kauffman sings his song loud and long, from the origins of life to the emergence of a global civilization. It may not be a song that everybody wishes to hear, and it may be a song with many clashing harmonies and unscheduled pauses — but I sure hope he goes on singing it. And I guarantee that any reader whose imagination has survived an academic education — or has never been exposed to one — will learn a lot, and be changed forever. □

Ian Stewart is at the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

Did the Romans exploit uranium?

Michael Vickers

Uranium Glass. By Ken Tomabechi. Iwanami Book Service Center, 2-3 Jinbocho Kanda Chiyoda-ku, Tokyo 101, Japan: 1995. Pp. 180. ¥7,000.

KEN Tomabechi has spent his professional life in the Japanese atomic energy industry, but for many years his hobby has been the collection and study of objects made from a material known variously as Vase-



Taken from Uranium Glass

Uranium glass vase made by the famous Moser Glassworks in Czechoslovakia (c.1920).

line or uranium glass in English, *Annagelb* or *Annagrün* in German, and *verre canari* in French. Vases, candlesticks, ashtrays, clocks, drinking vessels and even jewellery were fashioned in this curious medium (which is fluorescent under ultraviolet light) between the early nineteenth century and the Second World War, when uranium production took another direction. Uranium glass today is manufactured only for specific industrial rather than decorative purposes. Tomabechi describes well the peculiar scientific properties of uranium glass, and gives a comprehensive account of the centres at which it was produced in Europe, the United States and Japan. The resplendent colour illustrations show a wide range of uranium glass products from the author's collection.

There is, however, one section of the book that I find difficult to accept. This is where it is asserted that the Romans exploited uranium glass in order to achieve a light green tint in mosaics. The evidence is said to lie in some mosaic *tesserae* from a villa at Cap Posillipo destroyed at the same time as Pompeii in AD 79, supposedly brought back to Oxford by the excavator R. T. Gunther in 1911.

The relevant scientific literature certainly states as much. The reality, as described to me by David Brown, my predecessor as curator of the Roman section of the Ashmolean Museum, is somewhat different.

Uranium glass first entered the picture in 1911 when it was the only material with which the Daubeny Laboratory at Oxford could replicate the effect of the Roman glass. This was confirmed in the later 1960s when, at Franz Kirchheimer's suggestion, samples of fused glass preserved in test-tubes at the Ashmolean were analysed by the Max Planck Institute in Munich. None proved to be older than 1911, and none of the complete *tesserae* kept with them was found to contain uranium at all. The site at Cap Posillipo was also scanned with Geiger counters, and nothing of interest was registered. These researches were never published; Kirchheimer's interest seems to have waned once the Romans, and the possibility of 2,000-year-old uranium, fell out of the picture. To complicate matters further, the samples in Oxford were accidentally thrown away a few years ago, to the great disappointment of researchers who periodically come looking for them.

This is, however, simply intended to put the record straight, and in no way to belittle Tomabechi's achievement in writing the first book on a material of such scientific, social and aesthetic interest. □

Michael Vickers is at the Ashmolean Museum of Art and Archaeology, University of Oxford, Oxford OX1 2PH, UK.

Sensible to feeling

Stuart Sutherland

Emotional Intelligence: Why It Can Matter More Than IQ. By Daniel Goleman. Bantam: 1995. Pp. 342. \$23.95. To be published in the United Kingdom by Bloomsbury on 10 January at £16.99.

LIKE the Bible, the literature of social psychology can provide support for any view one happens to hold: it is ridden with contradictory results and pitfalls in interpretation. For example, it is easy to assume that because parents have been beastly to a child, he or she will turn into a beastly adult, but there are two alternative explanations. First, the parents may have treated the child badly because it was fractious from an early age (and continued to be so in later life). Second, the transmission of bad behaviour may have occurred through genetic inheritance affecting both parent and child in the same way, rather than through the child's environment.

Although Daniel Goleman does not consider these alternatives, he puts forward a case for the influence of parents on

personality, but he does so by selecting only studies that support his argument: there are many others that do not. For example, in a massive longitudinal study, George Vaillant found that an unhappy childhood was associated with success in later life. Goleman maintains, however, that "emotional intelligence" is an important factor in determining success. These curious buzz words mean the capacity to know and to verbalize what emotion one is feeling at a given time and why one is feeling it, the capacity to sense what emotion others are feeling and to empathize with them, and the ability to control one's reactions when emotionally aroused. But Goleman cites no evidence that these attributes of emotional intelligence actually occur together: indeed, cold people are usually restrained.

In an ingenious experiment, Walter Mischel presented 4-year-old children with a marshmallow each, telling them that he was going on an errand and that on return he would give them another one, provided they had not already eaten the first. When they were tracked down at the age of 18, it was found that those able to resist temptation had higher examination scores than those who could not. Moreover, their success at 18 was better predicted by the marshmallow test than by conventional IQ tests, although Goleman admits that IQ is a better predictor from about the age of 8 onwards. Indeed, emotional intelligence is not the only route to success: neither Hitler nor Stalin were notable for empathy or self-restraint, but by their own lights both were successful.

Where psychology leads, can neurophysiology be far behind? Goleman mentions Antonio Damasio's exciting work showing that signals associated with emotion that arise in the prefrontal lobes are important for making decisions. He also cites evidence that visual stimuli are analysed not merely by the visual cortex but also by a fast route through the amygdala, which enables people to respond quickly to threat, either with fight or flight: the same applies to other modalities. It is the longer pathways by way of the sensory cortices that enable spontaneous outbursts of emotion to be overcome. Interestingly, lesions in the left prefrontal lobes produce anger and fear, whereas lesions in the right can give rise to extreme cheerfulness. All this may be true, but it is worth observing that despite our anatomical knowledge of the brain's pathways, nobody can explain how their working leads to the subjective emotions of fear or anger and their accompanying reactions: the sympathetic system is aroused in similar ways in both.

Perhaps the most interesting part of *Emotional Intelligence* is the description of the programmes devised to teach that skill, which now exist in about 20 US schools, including ones for deprived

children. The pupils' interactions with one another are observed and they are taught on the spot to identify their own feelings and those of others, to restrain themselves, to delay gratification and to reduce stress: indeed, the only desirable quality on which no instruction is provided is a sense of humour, but that may be lacking in their earnest teachers. Outcome studies on these schools have apparently all been favourable, but Goleman does not give enough information to make it possible to judge them. Were the results significant? How long did the effects of the teaching last? Were the assessors blind to the treatment the children had received? Were there placebo groups to control for the Hawthorne

effect — the tendency for any intervention to improve matters?

Despite these omissions and despite his selective use of evidence, Goleman has written an interesting if somewhat naive book. He has done his homework well, describing many recent studies, although he ignores older work such as that of Stanley Schachter. Some will find the journalistic tone repugnant, but anyone interested in emotion is likely to discover challenging new ideas in this book: it should be read by social workers and agony aunts everywhere. □

Stuart Sutherland is at the Laboratory for Experimental Psychology, University of Sussex, Brighton BN1 9QG, UK.

Geological controversies hammered out

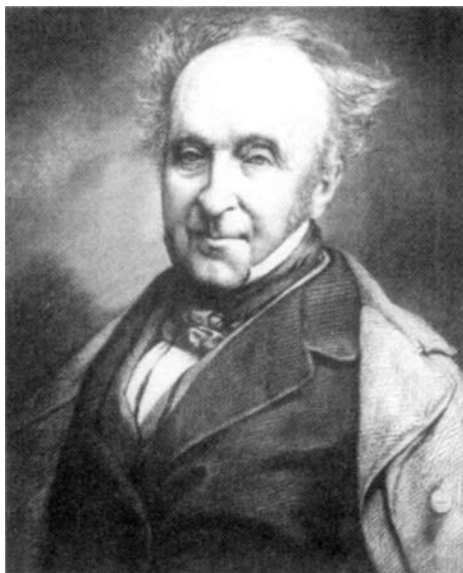
Nicolaas Rupke

Murchison in Moray: A Geologist on Home Ground. With the Correspondence of Roderick Impey Murchison and the Rev. Dr George Gordon of Birnie. By Michael Collie and John Diemer. *American Philosophical Society*: 1995. Pp. 263. \$20 (pbk).

ONE of the great and intricate constructs of modern science is the stratigraphic table, which comprises the time-rock units of geological history. The table acquired its nearly definitive form during the first half of the nineteenth century. Among the leading stratigraphers of the period was Roderick Impey Murchison (1792–1871), whose most famous publication, *The Silurian System* (1839; republished under the title *Siluria* in 1854, 1859 and 1867), was an important contribution to the construction of the Palaeozoic era (and erathem). Part of the fieldwork for this book was carried out in the north of Scotland, along the Moray Firth, north-east of Inverness. Here, in the Elgin area, a controversy developed over the stratigraphic position of rocks, which Murchison believed to be old red sandstone, but which proved to be new red (Permian-Triassic). One source of Murchison's information, in particular for the fourth edition of his great book, was a local clergyman-naturalist, George Gordon, who had a thorough knowledge of the geological structure of the region. The Murchison–Gordon correspondence, from 1858 to 1867, forms the second part of *Murchison in Moray*. In editing these letters, Michael Collie and John Diemer have made a valuable contribution to Murchison scholarship.

Murchison was, however, far more than a leading stratigrapher. He was a member of the magic circle of gentlemen-naturalists who ruled British science dur-

ing the Victorian era; he became director-general of the British Geological Survey; he carried out expeditions to Russia; and he was one of the founders of the Geographical Society, and enthusiastically supported explorations and surveys in the context of British colonial expansion. The imperial dimension of Murchison's activities has been discussed by James Secord and, in more detail, by Robert Stafford, who entitled his biography of Murchison *Scientist of Empire* (1989). Moreover, as Martin Rudwick long ago pointed out,



Roderick Impey Murchison: scientist of Empire.

Murchison, who began his career as a military man, had a disposition to turning scientific controversies into "paramilitary campaigns" against his adversaries, such as T. H. de la Beche (over the delineation of the Devonian system) and Adam Sedgwick (over the Cambrian/Silurian).

Such assessments of their hero are anathemas to Collie and Diemer, and the

first part of their book is designed to present a very different Murchison. To them, interpreting Murchison in the context of his career objectives and of the political dynamics of his metropolitan social circle is nothing less than sully the character and reputation of an upright Scotsman and an honourable geologist. The authors sternly censure "the ease with which some university-based critics stigmatize and condemn Murchison, sneering at him whenever the opportunity presents itself". In places, Collie and Diemer get hot under the collar defending Murchison against both his Victorian and present-day detractors. The latter are rapped over the knuckles for being "retrospective myth-makers". Stafford is the blackest of their *bêtes noires*, and one of his complaints about Murchison, namely that he engaged in élitist socializing in his London home, is countered as follows: "Actually, the criticism of Murchison's establishment in Belgrave Square was John Ruskin's, a person whose egotism and mean-mindedness were amply supported by a private income derived from the sale of sherry. Perhaps not enough sherry was consumed at a Murchison soiree."

The proper historiographical treatment of Murchison requires, in the view of the authors, stepping into his shoes, retracing his footsteps in the field and facing the technical geological problems he encountered. To this end, they use, in addition to Murchison's publications, the Gordon correspondence and also his notebooks —

the latter in preference to the journal he wrote towards the end of his life, criticizing along the way David Oldroyd for having given pride of place, in *The Highlands Controversy* (1990), to the journal rather than the notebooks.

There is no doubt that the authors have done a fine and valuable job in following Murchison into the Moray Firth region. Yet their historiographical approach is atavistic, carrying us back to a time when the history of science was for the larger part written by failed or retired scientists. Although their spirited attack on recent Murchison scholarship adds spice to the narrative, the authors might have done better to look for ways of integrating their solid, detailed work with the wider-ranging efforts of previous Murchison studies. The possibility of such an integration is indicated by their own conclusion: "Thus what seemed to be a straightforward argument about the interpretation of geological and palaeontological evidence was in fact ideologically complex and interesting." □

Nicolaas Rupke is at the Institut für Geschichte der Medizin, Georg-August-Universität, Humboldtallee 36, D-37073 Göttingen, Germany.

The end of history

Fred S. Rosen

Species and Specificity: An Interpretation of the History of Immunology. By Pauline M. H. Mazumdar. Cambridge University Press: 1995. Pp. 457. £40, \$64.95.

Immunology: The Making of a Modern Science. By R. B. Gallagher, Jean Gilder, G. J. V. Nossal and G. Salvatore. Academic: 1995. Pp. 246. £27.50, \$39.95.

WHEN Francis Fukuyama announced the end of history with the triumph of liberal democracy a few years back, he would have done better to confine his thoughts to immunology. Having solved the problem of the molecular basis of specificity, immunology reached the end of its history; it started out a century ago as a scientific endeavour to reach an understanding of and solution to many problems, but primarily an explanation of the molecular basis of immunological specificity. The field has reached a state of such maturity that the time has come to seek out its intellectual roots. That search has largely concentrated on German biological science in the mid- to late-nineteenth century. In 1991 A. I. Tauber and L. Chernyak wrote about the origins of Metchnikoff's science in the experimental embryology of Pander, Baer, Rathke and others, curiously scientists who worked primarily at the German-speaking universities in the Baltic rim. The Saturday afternoon in December 1882 when Metchnikoff discovered phagocytosis was not an epiphany in Messina. It came from decades of contemplating the function of mesodermal cells in invertebrates.

Now Pauline Mazumdar, a science historian, has made a scholarly foray into the origins of the concepts of immunological specificity which has taken her back to the German botanists and microbiologists wrestling with problems of speciation during the second half of the nineteenth century. They were divided between the pluralists and the unitarians, and their controversies came to affect the polemics of immunology. The visit of the country physician Robert Koch to the Institute of Plant Physiology, directed by Ferdinand Cohn, and the Institute of Pathology, under the direction of Julius Cohnheim, both in Breslau, provided a decisive moment in 1876 when Koch demonstrated the life cycle of the anthrax bacillus, from spore to pathogen — a single species. (He certainly received a better reception than his predecessor of almost a century earlier, another country doctor, Edward Jenner, whose reports the Royal

Society declined to publish.) Koch's demonstration was a triumph for the pluralists who had the faith that speciation could be determined in microorganisms. Paul Ehrlich was a direct heir of this pluralist tradition and it came to dominate the thinking of early immunologists.

Karl Landsteiner was the principal opponent of pluralist thinking in immunology, although he had plenty of support from Buchner, Bordet and others. *Species and Specificity* is largely a biography of this singularly brilliant and unattractive man. His mother, Franziska Hess Landsteiner, had a remarkable physical resemblance to another contemporary Moravian Jewess and ambitious parent, Amalia Freud. (The book is abundantly and well illustrated with many photographs.) Alas, Franziska took her young

Landsteiner and Weiner considered the Rh antigens to be different versions of a protein encoded by a single gene. Fisher, Race and Sanger, subscribing to the Batesonian view of genetics, incisively offered another explanation, that the Rh antigens were encoded by three linked alleles. This was the culmination of the philosophical arguments that had driven immunologists for half a century.

Immunology was born in conflict, and early on resorted to a bellicose vocabulary to depict its imagery. Reaching for military metaphors and similes, immunologists always talk about struggle, invaders, defences and so on. Immunology was also born amid the great triumph of synthetic organic chemistry. The chemists and their heirs, the serologists, dominated the field for more than half a century as the biological thinking of Metchnikoff was

swamped by his opponents. It was not until Burnet and Talmage began after the Second World War to frame the immune response in biological terms that stunning progress was brought about. The half-baked ideas about immunological specificity that dominated the early workers in the field can only be attributed to their forbidding ignorance. The technology of science was not prepared at that time to answer the vexing questions posed by, say, the neutralization of toxins by antitoxins. The controversy surrounding the Danysz phenomenon is a good example. When a toxin is added all at once to its antitoxin, the toxin is neutralized. When the toxin is added in aliquots every 30 minutes, the toxin is not completely neutralized. Bordet mistakenly compared this to dyeing filter paper, which would be sequentially less dyed if pieces of the paper were dipped sequentially into a dye

whereas a single piece of filter paper would be uniformly dyed at one exposure. Alas, they did not understand the reversibility of antigen-antibody interactions and the concept of equilibration.

Gallagher and co-editors have put together a compendium of charming reminiscences by many of the outstanding immunologists alive today. (Unfortunately there are some serious omissions, but the editors did not intend the coverage to be complete; it is a pity though that there are no women contributors other than Eva Klein, who writes a chapter with her husband George.) These short, concise autobiographical notes also mark the end of history. If you believe that great science, like great art, is divorced from the persona of its creator, don't read this book. □

Fred S. Rosen is at the Center for Blood Research, 800 Huntington Avenue, Boston, Massachusetts 02115, USA.

IMAGE
UNAMAGABLE
FOR A CORNABET
FOR REASONS
REASONS

Karl Landsteiner: principal opponent of pluralist thinking in immunology.

Karl to the Schottenkirche to be baptized lest the institutionalized anti-Semitism of the Hapsburg dominions block his road to success. It didn't help. Emaciated and half starved in the aftermath of the collapse of the Austro-Hungarian Empire, and certainly unrecognized for his brilliant discovery of the blood groups, Landsteiner took a position in a small charity hospital at The Hague where he had little time for research. He was literally rescued by Simon Flexner who took him to the Rockefeller Institute in New York where he spent the rest of his productive career, spanning two decades, including a Nobel prize in 1930. An authoritarian character himself, Landsteiner bitterly resented the authority of his saviour, Flexner. The book ends with an interesting exegesis on the rhesus (Rh) controversy. The discovery of the Rh blood groups by Weiner and Landsteiner led to a bitter and unresolved controversy. The unitarian view held by

The crystal structure of the GroES co-chaperonin at 2.8 Å resolution

John F. Hunt, Arthur J. Weaver*, Samuel J. Landry*†, Lila Gierasch*† & Johann Deisenhofer

Howard Hughes Medical Institute and Departments of Biochemistry and † Pharmacology, The University of Texas Southwestern Medical Center, Dallas, Texas 75235, USA

The GroES heptamer forms a dome, approximately 75 Å in diameter and 30 Å high, with an 8 Å orifice in the centre of its roof. The 'mobile loop' segment, previously identified as a GroEL binding determinant, is disordered in the crystal structure in six subunits; the single well-ordered copy extends from the bottom outer rim of the GroES dome, suggesting that the cavity within the dome is continuous with the polypeptide binding chamber of GroEL in the chaperonin complex.

CHAPERONINS are proteins that control, and under some circumstances catalyse, the folding of other proteins^{1,2}. The most thoroughly studied chaperonin system is the GroEL/GroES complex from *Escherichia coli*³, but, despite extensive investigation, the mechanism of GroEL/GroES-assisted protein folding remains obscure^{4,5}.

The crystal structure of the GroEL tetradecamer⁶ shows its subunits of relative molecular mass 60,000 (M_r 60K) to be arranged as a dimer of heptamers to form a cylinder with nearly proper seven-fold symmetry along its axis and two-fold symmetry perpendicular to its axis. Non-native substrate polypeptides bind to the end of the GroEL cylinder in the opening of its predominantly hollow core^{7–9}; chemical sequestration of these aggregation-prone species is a major function of GroEL^{1,2,4,5}. The GroES oligomer comprises a heptamer of 10K subunits. Electron micrograph studies^{7,8,10} show that GroES is positioned like a cap over the top of one of the two polypeptide binding chambers of GroEL in the stable GroEL₁₄–GroES₇–ADP₇ complex^{11,12}. The binding of GroES to GroEL is mediated at least in part by the 'mobile loop' of GroES, a 16-residue segment in each subunit that is disordered in the free heptamer but becomes immobilized when it binds to GroEL¹³. ATP hydrolysis by GroEL accelerates the release of the bound substrate polypeptide^{11,14}. When solution conditions support spontaneous folding in the absence of the chaperonin, ATP-induced release can lead to successful folding of the substrate polypeptide in the absence of GroES^{15,16}; otherwise, GroES is strictly required for successful chaperonin-assisted folding^{1,16}.

Two fundamental questions about the mechanism of GroEL/GroES-assisted protein folding remain unanswered. The first is whether the primary function of the chaperonin complex is to fold^{5,15,17} or to unfold^{12,18–20} non-native polypeptides. The 'Anfinsen cage' hypothesis maintains that the function of the chaperonin complex is to provide a protected environment, potentially the polypeptide binding chamber under GroES in the GroEL/GroES complex, where folding to the native state can be promoted without interference from the great reservoir of aggregation-prone species in the cell^{5,17}. Conversely, the 'iterative annealing' hypothesis¹² maintains that the function of the chaperonin complex is to partly unfold misfolded proteins^{12,18–20} (that is, kinetically trapped intermediates); some of the energy of ATP hydrolysis could be channelled into the conformational energy of the sub-

strate polypeptide, forcing it into a higher energy state from which it could once again attempt to fold correctly after being released to solution.

The second fundamental question concerns whether the productive protein folding reaction occurs with GroES and the substrate polypeptide on the same side (*cis*) or on opposite sides (*trans*) of the GroEL double cylinder^{7,12,21}. The substrate ornithine transcarbamylase is productively released and folded, starting from a sequestered position beneath GroES in the binding cavity of GroEL⁵¹, while the substrate pre- β -lactamase can be chemically crosslinked to GroES in the GroEL/GroES complex²¹. Although these observations support the *cis* mechanism, electron micrograph analysis shows the substrate malate dehydrogenase predominantly in the *trans* position during folding⁷.

With these basic questions unresolved, the exact role of the co-chaperonin GroES in the ATP-dependent protein-folding reaction remains uncertain. GroES increases the cooperativity of ATP hydrolysis by GroEL^{22,23}, leading to 'quantized' hydrolysis of all seven ATP molecules in one of the half-cylinders of GroEL¹². However, it is not clear whether the co-chaperonin plays any role in assisted folding beyond allosteric regulation of GroEL. In a *trans* complex the role of GroES must be limited to allosteric regulation, but in a *cis* complex GroES could participate directly in the protein folding reaction because it will be in immediate proximity to the substrate. In this case, GroES will form part of the surface of the chemically insulated folding compartment if the Anfinsen cage hypothesis is correct; alternatively, it could play a functional role in the unfolding reaction if the iterative annealing hypothesis is correct.

Here we present the atomic structure of the GroES heptamer. The structure that we observe is potentially highly flexible, inconsistent with our expectations for a molecule with the sole function of locking GroEL into a specific allosteric state. Other features in the structure suggest a possible functional role for the GroES co-chaperonin in the GroEL/GroES-assisted protein-folding reaction.

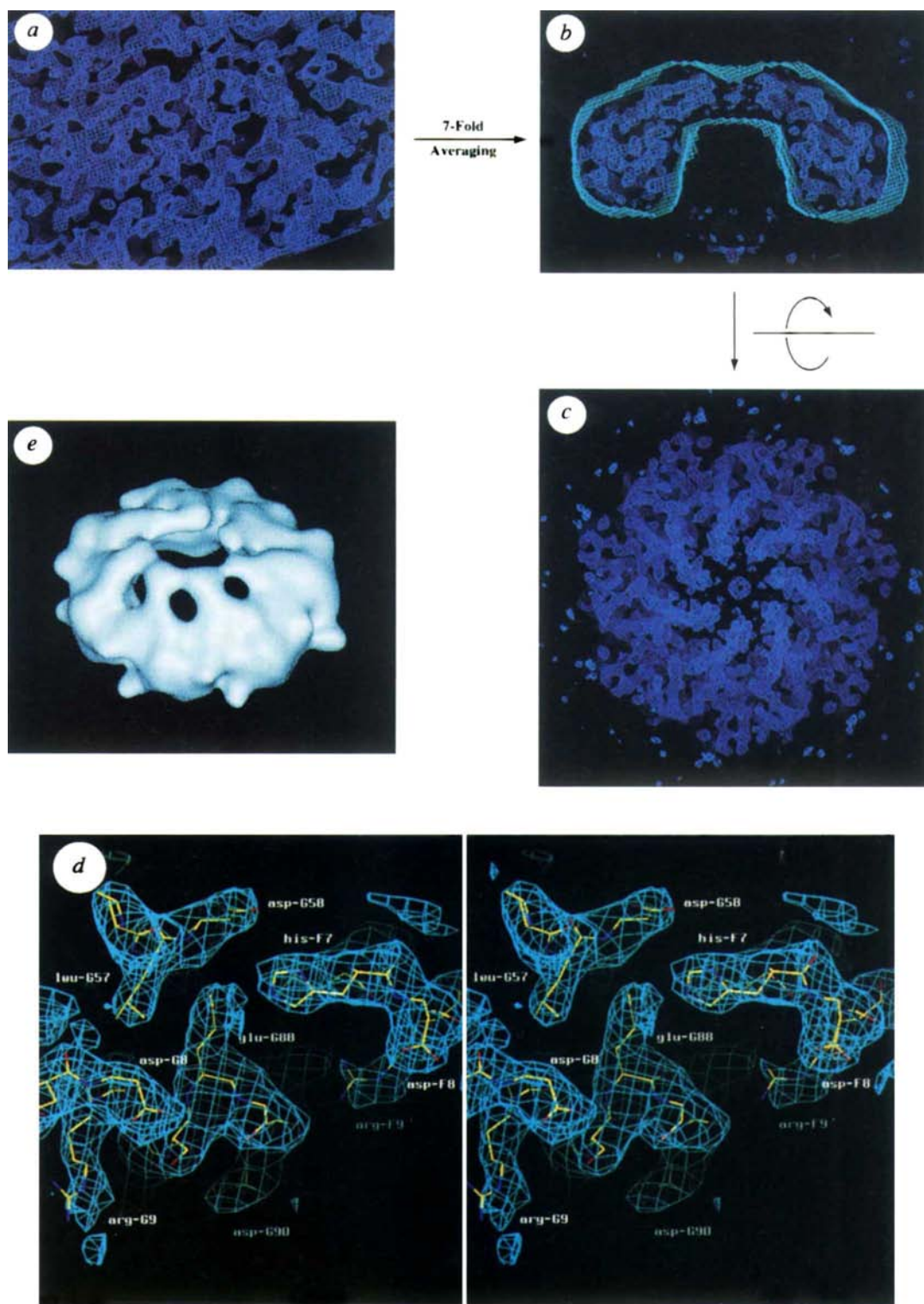
Structure determination

Crystals are grown in the orthorhombic space group $P2_12_12_1$ with one heptamer per asymmetric unit (Table 1a). We identified two heavy-metal derivatives with reasonable phasing power at low resolution but no useful phasing power beyond 4.5 Å (Table 1b, c). The Yb(OOCCH₃)₃ derivative showed one well-occupied but spatially diffuse metal site per asymmetric unit, and the CH₃HgCl derivative showed one poorly occupied metal site per asymmetric unit. The resulting multiple isomorphous replacement

*Present addresses: hkl Research Inc., 733 Seventh Avenue SW, Rochester, Minnesota 55902, USA (A.J.W.); Department of Biochemistry, Tulane University School of Medicine, New Orleans, Louisiana 70112, USA (S.J.L.); Department of Chemistry, University of Massachusetts, Amherst, Massachusetts 01003, USA (L.G.).

FIG. 1 Experimental electron-density maps. *a*, MIR map from MLPHARE^{32,40}. *b, c*, Map after proper seven-fold NCS averaging. *b*, The map has the same orientation and section as the MIR map (*a*); the molecular envelope used for averaging is superimposed. *c*, The map is in projection after a 90° rotation relative to the view shown in *b*. *d*, Map at 2.8-Å resolution after improper NCS averaging starting with combined phases from proper averaging and a poly-alanine partial model. A region along the inside wall of the GroES dome is shown in an orientation analogous to that in *b*. *e*, Surface rendering of a $(2F_o - F_c)$ map from the refined model filtered to low resolution using the Wang-Leslie procedure^{45,46} with an averaging radius of 8 Å.

METHODS. All procedures related to density averaging were performed using the DEMON/ANGEL suite of programs²⁴. In the initial proper averaging procedure (*b, c*), an identical 51.43° rotation around the central symmetry axis was assumed to relate each subunit to its neighbour. The location and orientation of the proper symmetry axis were optimized by maximizing the linear correlation coefficient before and after seven-fold averaging within a concentric cylindrical envelope. We derived a molecular envelope for the heptamer by applying the Wang-Leslie algorithm^{45,46} after seven-fold averaging in a pseudo-P1 unit cell centred on the putative molecular aggregate²⁴. The resulting envelope was scattered with small islands of occupied points, but contained a reasonable envelope for the central molecular aggregate. The envelope for this aggregate was isolated using CNCTDENV²⁴, which extracts the largest connected set of occupied points, and was then expanded by adding a single layer of points to its surface. The improper NCS averaging procedure (*d*) used 42 different symmetry operators to represent the relationship between each individual subunit and its six neighbours. The operators were deduced by optimizing the superposition of the polypeptide backbone of the subunits in a partly refined model, and a molecular envelope was derived individually for each subunit based on the complete atomic model using MAMA⁴⁷. Both averaging procedures effected phase extension from 3.5 to 2.8 Å using standard protocols²⁴. To produce the low-resolution map (*e*), a sigma-a weighted $(2F_o - F_c)$ map



was calculated at 2.8 Å resolution in the true unit cell and then a volume comprising the heptamer plus a boundary region of 10 Å was isolated and treated as an artificial P1 unit cell. A solvent mask was applied in real space to flatten the electron density from the crystallographic sym-mates using a modified version of AVERG²⁴, and a set of pseudo-structure factors were calculated by Fourier inversion. These structure factors were filtered to low resolution using the Wang-Leslie algorithm^{45,46} before Fourier synthesis. This figure was produced using O⁴³, except *e*, which was produced using AVS.

(MIR) map (Fig. 1*a*) could not be interpreted in molecular detail but was adequate to allow solution of the structure using seven-fold non-crystallographic symmetry (NCS) averaging²⁴.

The self-rotation function provided very weak evidence of a seven-fold symmetry axis orientated close to the crystallographic *c*-axis^{25,26}. Visual inspection of the MIR map revealed weak seven-fold symmetry around an axis inclined approximately 20° from the crystallographic *c*-axis. After optimization of the orientation and position of this axis using a real-space search, proper seven-fold NCS averaging was performed inside a single molecular envelope encompassing the heptamer (Fig. 1*a,b*). Substantial deviation of the exact non-crystallographic symmetry from the assumed proper symmetry caused the *R*-factor for *F_o* versus *F_c* to remain at approximately 32% throughout this procedure. Nonetheless, it yielded a map in which the GroES monomer could be interpreted in terms of a polyaniline model comprising 64 residues in seven segments.

Phase combination between the polyaniline model and the averaged phases yielded a map in which the chain could be traced for the entire monomer with a single interruption at the mobile loop. We assigned the sequence using this map based on the connectivity of the backbone combined with anomalous diffraction from Met 86 and well-defined side-chain density for a limited number of residues. To verify the sequence assignment, improper seven-fold NCS averaging was performed, starting with the combined phases from the initial polyaniline model (a side-chain-free model). This procedure yielded an *R*-factor for *F_o* versus *F_c* of 13% and a map with well-defined side-chain density for the vast majority of the residues in the monomer (Fig. 1*d*). The current atomic model comprises 80 residues each for six subunits, and all 97 residues for the seventh subunit in the heptamer (Table 1*d*).

Structure of the GroES monomer

Each GroES subunit forms an anti-parallel β -barrel with a

TABLE 1 Summary of crystallographic data

Space group: $P2_12_12_1$		Unit cell parameters at 130 K: $80.3 \times 141.5 \times 55.1 \text{ \AA}$ (90° , 90° , 90°)				
(a) Data collection statistics for native and derivative crystals						
	Native	Yb(OAc) ₃			CH ₃ HgCl	
Resolution limit (Å)	2.7	2.7			2.7	
R_{sym} (%)	3.6	5.6			5.1	
Unique reflections	16,087	17,157			16,911	
Observations	74,936	110,440			79,344	
Completeness to 3.5 Å (%)	91	96			97	
Completeness to limit (%)	82	85			82	
(b) Heavy-metal site parameters						
	Real	Relative occupancy		Anisotropic B -factors		Isotropic B -factor
		Anomalous				
Yb(OAc) ₃ Site 1	1.00	0.83		130, 152, 349		
CH ₃ HgCl Site 1	0.16	0.16		16, 79, 99		
Site 2	0.05	0.04				68
(c) Heavy-metal phasing statistics versus resolution						
	11.0 Å	7.6 Å	5.8 Å	4.7 Å	4.0 Å	Overall (4 Å)
Yb(OAc) ₃ Mean $ \Delta F/F $	0.16	0.12	0.14	0.12	0.12	0.12
Phasing power	2.9	1.7	1.2	0.5	0.2	0.6
R_{Cullis} (centric)	0.39	0.57	0.67	0.89	0.96	0.68
CH ₃ HgCl Mean $ \Delta F/F $	0.06	0.05	0.07	0.06	0.07	0.06
Phasing power	0.7	1.2	1.2	0.8	0.5	0.8
R_{Cullis} (centric)	0.69	0.70	0.66	0.81	0.89	0.76
Mean figure of merit	0.79	0.71	0.55	0.36	0.23	0.39
(d) Refinement statistics						
			6.5–2.8 Å			17–2.8 Å
R_{cryst} (%)			23.5			26.1
R_{free} (%)			28.8			30.6
Bond length deviation (r.m.s.)			0.016 Å			
Bond angle deviation (r.m.s.)			2.0°			

Summary of crystallographic data. Standard definitions are used for all of the statistical parameters, as compiled previously³⁸. (a), All values were calculated using PROTEIN³⁹. The symmetry *R*-factors are based on intensity. Both the *R*-factors and the completeness were calculated using data with $F \geq 2\sigma_F$. The resolution limit is defined as the last shell in the data set for which there is at least one-third completeness. (b), (c), All values were calculated using MLPHARE^{32,40} except for $|\Delta F/F|$, which was calculated using PROTEIN³⁹. (d), Statistics are presented for the model comprising 80 residues for 6 of the 7 subunits and all 97 residues for the remaining subunit (without added H₂O); all values were calculated using X-PLOR⁴¹. After purification¹³, the GroES protein was transferred to 10 mM BES, pH 7.0, and concentrated to 25 mg ml⁻¹. Crystals were grown by sitting drop vapour diffusion at 21 °C; the protein solution was mixed 1:1 with 20% PEG-400, 10 mM KCl, 160 mM Mg(OOCCH₃)₂, pH 5.0, and equilibrated against a reservoir containing 2.5 M NaCl, 80 mM Mg(OOCCH₃)₃, pH 5.5. The crystals grew to approximately $350 \times 300 \times 150 \mu\text{m}$ in 4 days. They were stabilized in 40% PEG-400, 10 mM KCl, 200 mM Mg(OOCCH₃)₂, pH 5.5, for at least 4 h before flash freezing in liquid ethane or propane for data collection. (Freezing in liquid nitrogen produced daunting crystal-to-crystal non-isomorphism which was eliminated by freezing in liquid alkane.) Crystallographic data from a Xuong-Hamlin area detector were integrated and scaled using XDS and XSCALE⁴² and merged using AGROVATA³². The model was built in O⁴³ and refined using X-PLOR⁴¹ with Engh and Huber parameters⁴⁴. Individual vicinally restrained isotropic *B*-factors⁴⁵ have been refined for all atoms. However, a strong seven-fold NCS restraint⁴¹ was maintained on both the positions ($300 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$) and the *B*-factors ($1.0 \text{ \AA}^2$ target deviation) of all of the atoms in 54 residues and the backbone atoms in an additional residue in every subunit; the conformation of the remaining residues deviated from the non-crystallographic symmetry in at least one subunit. Including all data, the crystallographic *R*_{free} is 32.2% at 6.5–2.8 Å and 33.4% at 17–2.8 Å. To avoid model bias, the side chains of the conformationally sensitive residues Arg 47, Glu 50, and Lys 55 have been omitted from refinement by setting the occupancy to zero for atoms beyond C β . Excluding glycine and proline, 88% and 98% of the residues, respectively, are in most-favoured regions and allowed regions of the Ramachandran plot. His 7 accounts for 7 of the 8 residues with disallowed Φ , Ψ angles.

β -hairpin extending radially outward from the top (Fig. 2*a-d*). The β -barrel is topologically irregular because the antiparallel hydrogen-bonding pattern is broken between the two strands that close the wall of the barrel beneath the β -hairpin. An additional β -strand at the amino terminus of each monomer is incorporated into the β -barrel in the adjacent subunit (Figs 2*d* and 3*b*). The core of the β -barrel is strongly hydrophobic.

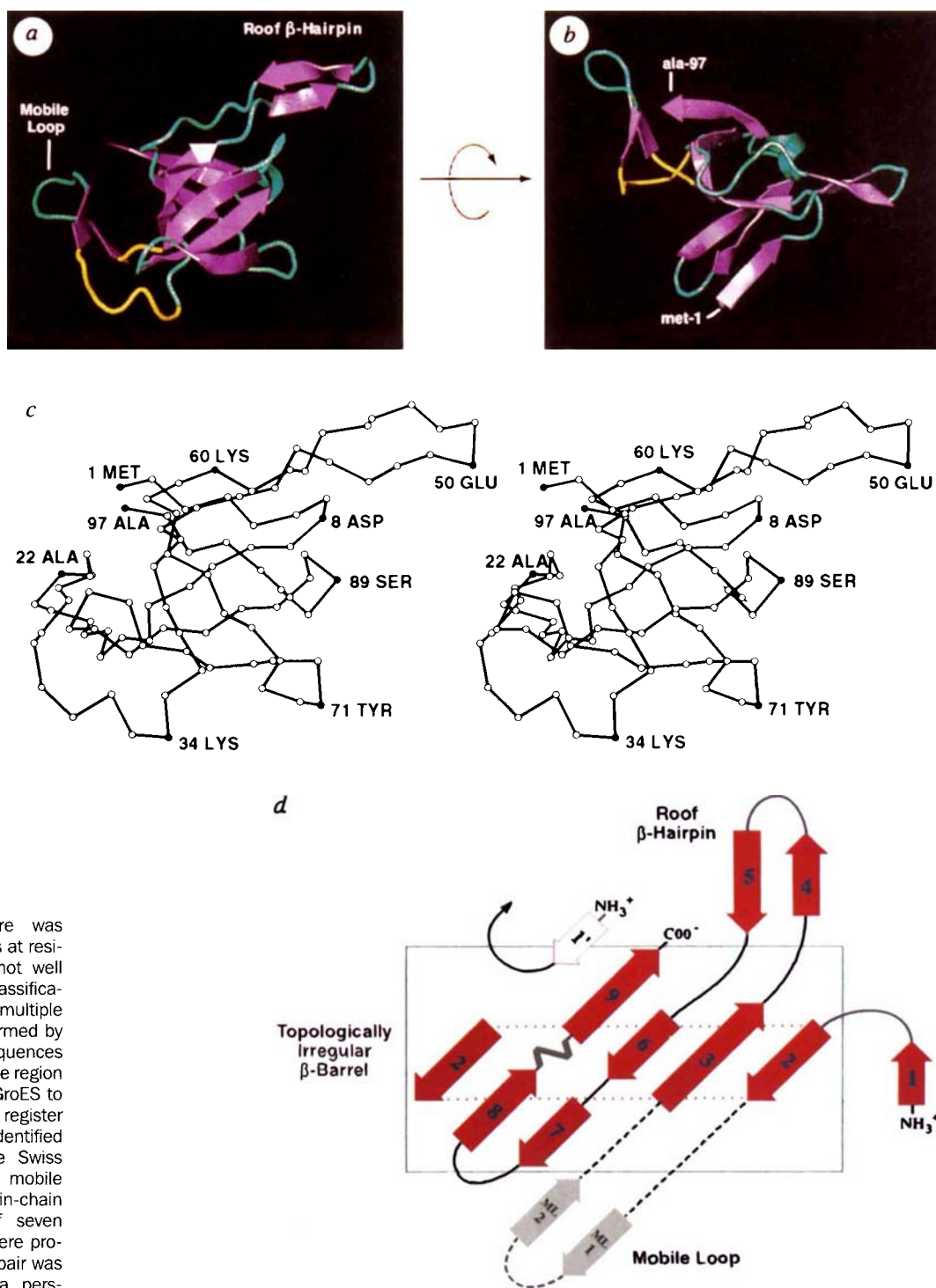
In six of the subunits, we observe electron density for residues Met 1 through Lys 15 and residues Ala 33 through Ala 97. By using nuclear magnetic resonance (NMR) spectroscopy, it has been shown that the polypeptide segment between residues Glu 16 and Ala 33 is disordered in uncomplexed GroES in solution¹³. We

observe electron density for this mobile loop segment in only one subunit in our structure (Fig. 2*a-c*). This single ordered copy forms a β -hairpin which is sandwiched between the wall of one GroES heptamer and the roof of another heptamer in a crystal packing contact. The mobile loop has previously been shown to be directly involved in ADP-dependent binding to GroEL¹³ and to adopt a different β -hairpin conformation in the complex compared to that observed in our crystal (S.J.L., unpublished data).

Figure 2*e* shows a phylogenetic alignment^{27,28} of GroES beneath a plot of the average main-chain *B*-factor in the crystallographic model. The β -strands in the protein coincide with the zones of minimum main-chain *B*-factor with the exception of strands

FIG. 2 Structure of the GroES monomer. *a, b*, Orthogonal views of a ribbon diagram of the single GroES subunit in which the mobile loop is observed. The residues linking the β -hairpin in the mobile loop to the core structure observed in the other six subunits are coloured yellow. *c*, C α stereo diagram of the same subunit. *d*, Topology diagram for the GroES subunit; the arrows represent β -strands, and the green zigzag represents a single turn of 3₁₀ helix. *e*, A phylogenetic alignment^{27,28} of proteins in the cpn10 family is shown beneath a schematic representation of the secondary structure of the monomer; the average main-chain *B* factors in the crystallographic model are plotted at the top. Positions where absolute sequence identity is observed throughout the phylogenetic tree are highlighted; positions where either the acidic, basic or aromatic character of the residue is preserved are outlined in a box. Residues forming the hydrophobic core of the GroES β -barrel are underlined. Average *B*-factors greater than 100 have been clipped, affecting six residues in surface loops. The β -turn types²⁹ are indicated beneath the structure schematic diagram.

METHODS. Secondary structure was assigned using DSSP⁴⁸. The turns at residues 20–24 and 78–81 are not well enough defined to allow reliable classification of their type. The automated multiple sequence alignment²⁷ was performed by PHD²⁸, although nine of the sequences had to be manually realigned in the region between Asp 69 and Glu 76 in GroES to bring their aromatic residues into register with Tyr 71; sequences are identified according to their code in the Swiss protein database. Except in the mobile loop (grey lines), the main-chain *B*-factors reflect the mean of seven subunits. The ribbon diagrams were produced using O⁴³, and the stereo pair was produced using PSFRODO (D. Xia, personal communication).



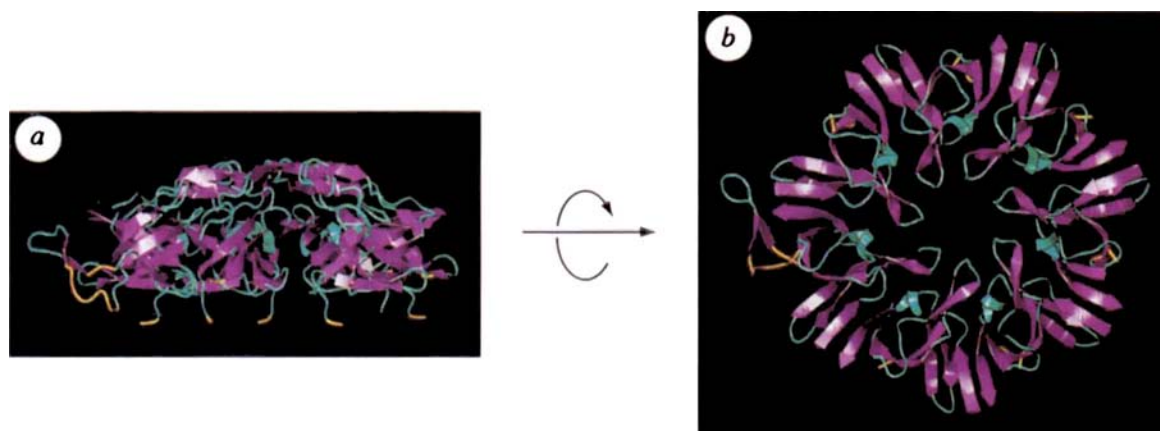


FIG. 3 Structure of the GroES heptamer. *a*, *b*, Orthogonal views of a ribbon diagram (produced using O). The segments of the polypeptide chain flanking the mobile loop are shown in yellow.

positioned like a cap over the top of the GroEL cylinder, the unoccupied volume within the GroES dome will be continuous with the central binding cavity in GroEL in the chaperonin complex.

Structural chemistry of the GroES dome

The evolutionarily conserved aromatic residue Tyr 71 is located on the inner wall of the GroES dome near its base, producing a hydrophobic environment in the orifice (Fig. 4*a*). Electron micrograph image reconstruction studies suggest that a substrate polypeptide bound to GroEL extends out of the top of the central cavity into the region that is eventually occupied by the base of the GroES dome in the chaperonin complex⁷. Therefore, the Tyr 71 ring of GroES is correctly positioned to interact with the substrate in a *cis* complex. A ring of solvent-exposed aromatic residues in GroEL constitute a critical determinant of its binding site⁹ for non-native polypeptides. In the crystal structure of GroES, the conformation of Tyr 71 varies in the different subunits, indicating a conformational flexibility that could be useful in binding structurally and conformationally heterogeneous substrate polypeptides.

There is an intense accumulation of negative charge around the orifice in the centre of the roof of the GroES dome (Fig. 4*a,c*). The glutamate residues at positions 50 and 53 in the monomer flank the β -turn located at the tip of the β -hairpin which forms the roof. This $\alpha_R\alpha_R$ β -turn²⁰ represents the closest point of approach of the polypeptide chain to the seven-fold symmetry axis of the heptamer. The resulting close apposition of 14 glutamic acid residues should produce significant coulombic repulsion between the seven β -hairpins forming the roof of the GroES dome. Indeed, an electrostatic calculation using the program GRASP³⁰ shows a deeply negative electrostatic potential near the centre of the roof (Fig. 4*c*).

Although the side chains at position Glu 53 are well ordered, all seven of the side chains at position Glu 50 are disordered in the crystal structure with no significant electron density in the final maps. Given the observed C α positions it would not be possible for the Glu 50 side chains from opposite sides of the heptamer to form tight ligands to a single cation in the centre of the roof orifice. In the crystal, the glutamate cluster could be stabilized by electrostatic shielding mediated by the 200 mM Mg²⁺, as well as by partial titration of the carboxylate groups at pH 5.5. Because of its low electron density, it would not be possible to observe differential accumulation of Mg²⁺ in our crystallographic maps. However, we see direct evidence of electrostatic shielding of the glutamate cluster by Yb³⁺ in the Yb(OOCCH₃)₃ heavy-metal derivative. Difference Fourier maps for this derivative are dominated by a single intense peak coinciding with the centre of the roof orifice.

This peak has an irregular shape and a diameter of approximately 8 Å, consistent with the very high *B*-factors determined for this site during heavy-metal parameter refinement (Table 1*b*). Therefore, although the lanthanide cation accumulates in this region of negative electrostatic potential, there is no evidence that it organizes a well-defined binding site in the heptamer, even at a concentration of 10 mM.

It is more difficult to explain how the glutamate cluster could be stabilized at physiological ionic strength and neutral pH. Given the observed backbone conformation, Arg 47 is the only residue close enough to Glu 50 to form a salt bridge. The side chain of Arg 47 is disordered in five subunits in the heptamer; in the remaining two subunits, it participates in a hydrogen-bonding array from Asp 8 to Arg 47 to Glu 88 to His 7 in the adjacent subunit (Fig. 4*d,e*). This conformation precludes interaction with Glu 50. However, a single 120° X₁ rotation brings the guanidino group of Arg 47 into a position where it could form a hydrogen bond with Glu 50 from the adjacent subunit (with both side chains in rotamer conformations). Although this interaction is not observed in the electron-density maps, it is the only candidate for direct electrostatic stabilization of the glutamate cluster in the roof orifice. Although Lys 55 (Fig. 4*e*) provides some counter charge for residues Glu 50 and Glu 53, it is not in a reasonable position to interact directly with either acidic group, and its presence is not adequate to mitigate the deeply negative electrostatic potential at the centre of the roof (Fig. 4*c*).

In this context, it was surprising to find that the packing interactions between the β -hairpins that form the roof of the GroES dome appear to be weak. On average, only 133 Å² of accessible surface area^{31,32} is buried per interface per β -hairpin. Moreover, very few inter-atomic contacts are observed between adjacent β -hairpins (an average of eight van der Waals contacts ranging from a minimum of one to a maximum of 13). A single hydrogen bond is observed in this interface between the side chain of Lys 55 and the backbone of Gly 52. However, the final electron-density maps indicate that Lys 55 is at least partly disordered in five subunits, and are consistent with a structural polymorphism in two subunits involving a mixture of the inter-subunit hydrogen bond to Gly 52 and an intra-subunit hydrogen bond to Asp 58 (Fig. 4*e*). For comparison, there is an average of 627 Å² of accessible surface area buried per subunit in the interface between adjacent β -barrels in the heptamer, with a mean of 95 van der Waals contacts and 11 hydrogen bonds. The weak packing density in the interface between the roof β -hairpins is also apparent in Fig. 1*e*, which shows the final electron-density map filtered to low resolution. Finally, the crystallographic model shows a local peak in the average main-chain *B*-factor coinciding with the roof β -hairpin (Fig. 2*e*), indicating enhanced thermal motion in this region of the protein.

FIG. 4 Structural chemistry of the GroES dome. *a*, Cut-away view of the cavity within the dome. The polypeptide backbone and the side chains of residues Glu 50, Glu 53 and Tyr 71 are shown for four subunits. Residue Glu 50 is located on the lower surface of the roof at the $(i + 1)$ position of the β -turn in the roof hairpin, and Glu 53 is located on the upper surface of the roof. *b*, The location of the phylogenetically conserved glycine residues in GroES. The tube representing the backbone is coloured yellow between the $C\alpha$ atoms of the two flanking residues. Gly 44 and Gly 46 form the long, yellow segment immediately preceding the lower, N-terminal strand of the roof hairpin. *c*, Surface map of the electrostatic potential of the GroES dome as calculated by GRASP³⁰. The blue and red colours represent positive and negative potential, respectively; a fully saturated colour indicates a potential of ≥ 5 kT (at physiological temperature). His 7 was assumed to carry a net charge of +1 in all subunits; the extent and intensity of the region of negative potential is increased substantially if the histidines are assumed to be neutral. *d, e*, Electrostatic interactions in the roof of the GroES dome. Segments from three subunits are displayed. Acidic side chains are shown in red, basic side chains in blue, and atoms from the phylogenetically conserved glycine residues at positions 44, 46 and 52 in yellow. The green atoms indicate positions that are observed to be at least partly disordered in the crystal structure. The side chain of residue Glu 50 is disordered in all subunits, and is shown in a rotamer conformation. The side chain of Arg 47 adopts the displayed conformation in two subunits but is disordered in the other subunits; this conformation makes hydrogen bonds to both Asp 8 and Glu 88. The side chain of Lys 55 adopts the displayed conformation in one or two subunits but is at least partly disordered in the other five; this conformation makes a hydrogen bond to the carbonyl group of Gly 52 in the roof β -hairpin of the adjacent subunit.

METHODS. O was used to produce *a* and *b*, MOLSCRIPT⁴⁹/RASTER3D⁵⁰ were used for *d* and *e*. The electrostatic calculation (*c*) used only full

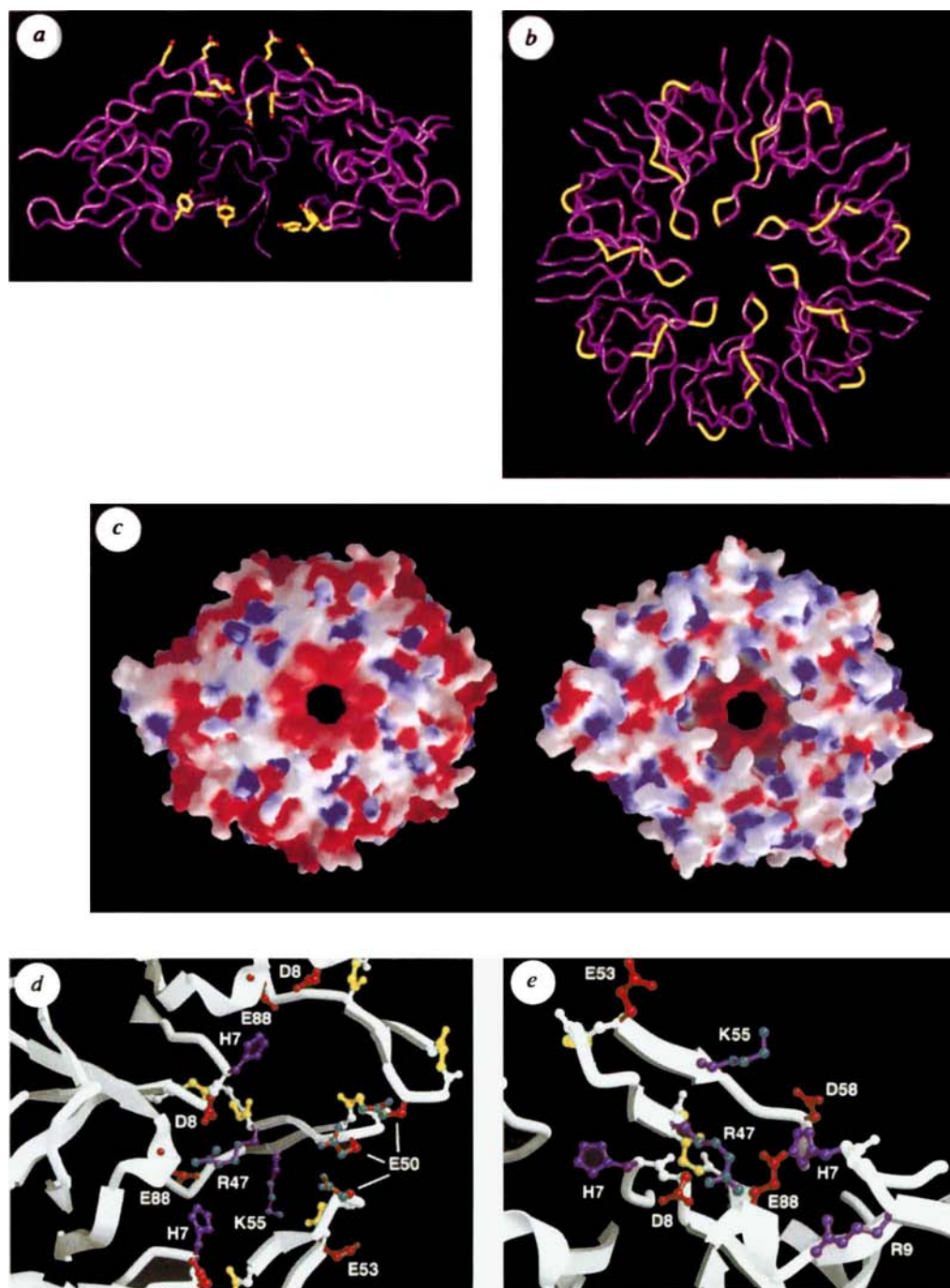
charges, but a qualitatively similar potential map was obtained after including partial charges; an ionic strength of 0.1 M was assumed.

The weak structural interactions coupled with the intense accumulation of negative charge led us to consider the possibility that the roof of the GroES dome might be a metastable structure. The four phylogenetically conserved glycine residues in GroES are all distributed around the roof β -hairpin (Fig. 4*b*). These residues are correctly positioned to provide structural flexibility for a conformational transition involving the roof. Moreover, backbone atoms from residues Gly 44 and Gly 46 form a molecular clamp that rotates the carbonyl group of His 7 into a disallowed Φ , Ψ value (Fig. 4*d, e*), a feature that is strongly correlated with enzymatically functional regions in protein molecules³³. (His 7 is the only residue in the 2.8-Å model with disallowed backbone dihedral angles, in spite of the fact that it

is in the region of the structure with the best-defined electron density (Fig. 1*d*) and the lowest main-chain *B*-factors (Fig. 2*e*).)

The subunit–subunit interface

Every subunit in the GroES heptamer is located in a different chemical environment in the crystal, producing a substantial deviation from proper seven-fold symmetry (Figs 1*e* and 5*a*). Taking the seven possible pairs of adjacent subunits and performing a least-squares superposition of one monomer reveals both the rigorous conservation of the core structure of the β -barrel and the highly irregular packing of the adjacent subunits (Fig. 5*a*). If the arrangement of the subunits obeyed proper symmetry, the second



subunit in each pair would also superimpose perfectly, even though the explicit superposition applied only to the first. Instead, a broad and irregular distribution is observed in the position of the second subunit, indicating a high degree of flexibility in the subunit–subunit interface. This structural plasticity is reflected in wide variations in the quantitative parameters of the seven interfaces (data not shown).

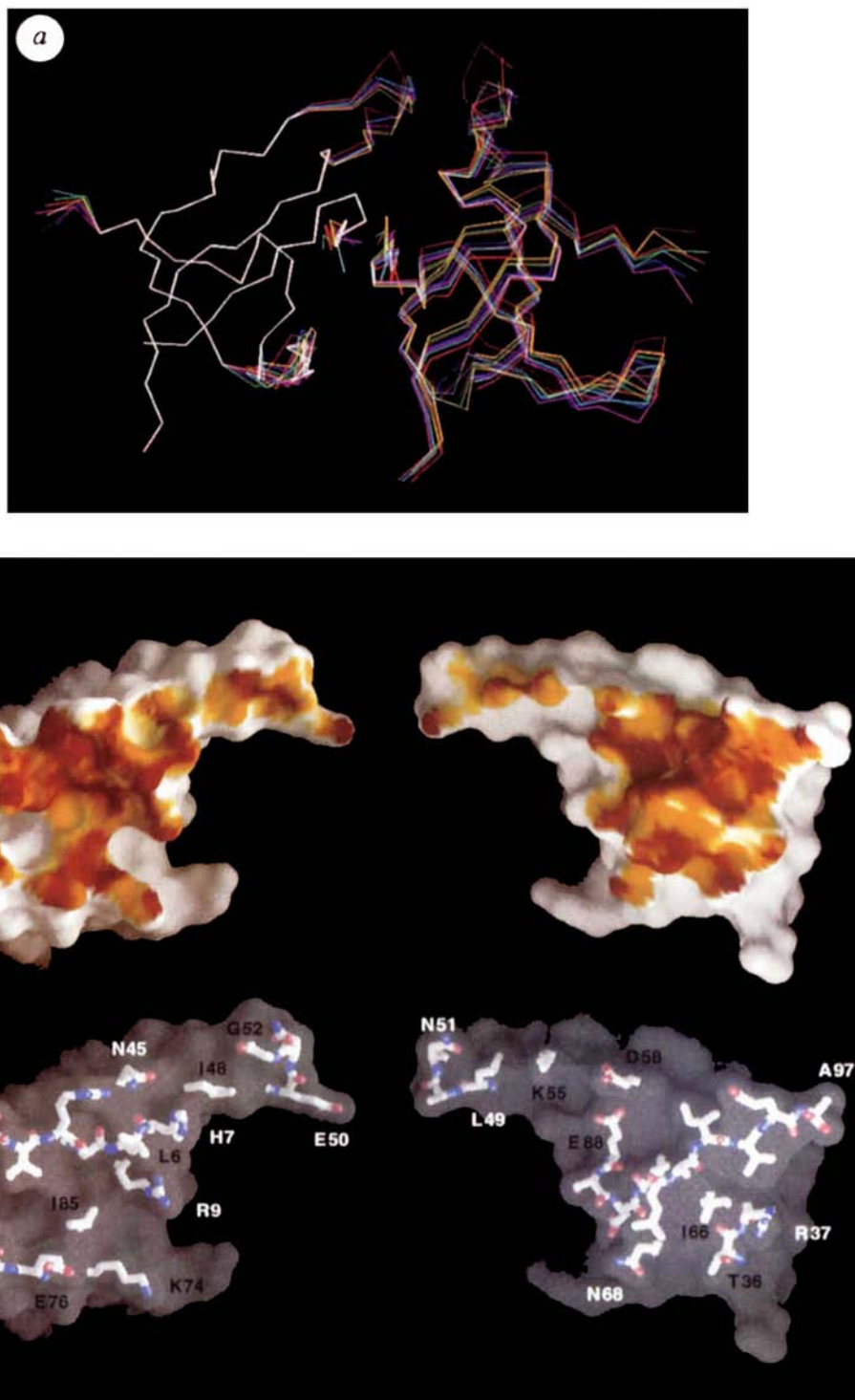
The mean value of the buried accessible surface area per subunit in the interface ($760 \text{ \AA}^2$) is similar to the values observed for oligomeric interfaces in other proteins of similar size³⁴. None the less, the GroES heptamer dissociates to monomers below a concentration of approximately $0.7 \mu\text{M}$ (in monomer)³⁵, corre-

sponding to a relatively weak binding energy of about -7 kcal mol^{-1} of interface.

Figure 5b shows the molecular surface of two interacting subunits in the heptamer^{30,36}. The dominant interaction in the interface is the anti-parallel zippering of the first β -strand in the subunit on the left to the final β -strand in the subunit on the right combined with a series of intimate and complementary interactions between the residues in the flanking β -turns. A second tier of interactions near the bottom of the subunits involves primarily the long side chains of Lys 74, Glu 76 and Arg 37 which stretch across the interface. The flexibility of these side chains and of the polypeptide backbone at residues 1–3, 76–79 and 96–97

FIG. 5 The structural basis of the plasticity of the subunit–subunit interface. **a**, All seven pairs of adjacent subunits in the heptamer ring were isolated and aligned based on a least-squares superposition of the NCS related $\text{C}\alpha$ atoms in the subunit on the left (that is, excluding all atoms from the subunit on the right). The $\text{C}\alpha$ chain trace for each pair of subunits is shown in a single colour, with seven different colours used for the seven pairs; the white colour of the chain in the subunit on the left results from the superposition of the seven colours. **b**, Surface complementarity and molecular structure of the subunit–subunit interface. Subunits D and E in the heptamer were rotated apart (around an axis parallel to the seven-fold axis) to reveal the details of their molecular interaction. **b**, Top, colour maps of the surface complementarity³⁶ in the interface; crimson encodes high complementarity, yellow encodes intermediate complementarity. **b**, Bottom, the surfaces have been rendered transparent to reveal the chemical constituents of the interface; only residues making van der Waals contact in the interface are shown.

METHODS. O^{43} was used to produce **a**, and was also used to perform the least-squares superposition of the subunits; the superposition was limited to the set of $\text{C}\alpha$ atoms to which the NCS symmetry restraint was applied during refinement. The surface complementarity was calculated using S_c ³⁶ with default parameters. GRASP³⁰ was used to produce **b**.



contributes to the plasticity of the interface. As a result, the subunits pivot around a core of more highly conserved interactions involving residues Pro 5 to Arg 9 and residues Ala 93 to Glu 88. The conformational flexibility of many of the chemical groups in the interface suggests that its formation may involve a significant loss of configurational entropy, which could account for its modest binding energy³⁷.

Possible functional implications

The conformational plasticity of the GroES dome combined with the potentially metastable structure of its roof suggest several possibilities concerning function. If the productive folding reaction involves a *cis* complex with the substrate polypeptide and GroES bound to the same ring of GroEL, a question arises as to how the substrate exits from the enclosed binding chamber. If the purpose of the chamber were to provide a sequestered environment for folding to the native state, it would seem reasonable for the native protein to diffuse away coincident with release of GroES. However, efficient release of non-native polypeptides from the complex has been demonstrated^{12,20}. Although it might be possible for a non-native polypeptide to diffuse away coincident with release of GroES, it must do so simultaneously with the return of the binding cavity to its high-affinity state for non-native polypeptide.

The observed structure of GroES suggests a speculative alternative: non-native polypeptide could exit from the GroEL binding chamber through the GroES ring. A conformational change resulting in the opening of the roof of the GroES dome would create a 30 Å orifice in the top of the binding chamber. This is slightly too small to pass the larger GroES-requiring substrates, but a compression in the GroEL binding cylinder during the ATPase cycle could push a substrate polypeptide in a molten globule state through the elastic orifice of GroES. The mechanical work done on the substrate polypeptide in pushing it through the limiting orifice of GroES would force it into a more extended conformational state from which refolding to the native state could proceed spontaneously, as suggested by the iterative anneal-

ing hypothesis^{12,18}. Another possibility is that the GroES ring itself could be played apart during some stage of the ATPase cycle of GroEL to enlarge the orifice and facilitate passage of the substrate polypeptide. Such a conformational change is plausible in the context of the relatively weak binding energy of the subunit-subunit interfaces in the GroES ring, which could potentially be weakened even further in the GroEL/GroES complex.

Although it is not possible to predict the occurrence of a specific protein conformational change, it is easy to foresee a structural mechanism by which the strength of the subunit-subunit interface could be modulated. The phylogenetically conserved basic residue Arg 9 makes an average of 11 van der Waals contacts and two hydrogen bonds in the subunit-subunit interface (Fig. 5b). The guanidino group of this residue is located 10 Å from the base of the GroES dome (Fig. 4e), putting it in position potentially to interact with GroEL. If Arg 9 were to be pulled down towards GroEL in the chaperonin complex, the interface between the subunits would certainly be weakened. The guanidino group of Arg 9 is located 5 Å from the imidazole group of His 7, a residue that participates in a series of structural interactions with the roof (Fig. 4d,e), as well as making an average of nine van der Waals contacts and one hydrogen bond in the subunit-subunit interface (Fig. 5b). The backbone of His 7 is held in a sterically strained conformation by a structural clamp formed between the phylogenetically conserved Gly 44 and Gly 46. In a subset of monomers in the crystal, His 7 participates in two different hydrogen-bonding arrays, both of which should be destabilizing to the roof: an array involving Asp 8, Arg 47 and Glu 88, which should weaken the electrostatic shielding by Arg 47 of the acidic cluster formed by Glu 50 in the roof orifice; and an array involving Lys 55 and Asp 58, which disrupts the hydrogen bond between Lys 55 and Gly 52 from adjacent roof β -hairpins (Fig. 4d,e). Thus electrostatic interaction between Arg 9 and His 7 could couple the movement of Arg 9 to other conformational rearrangements in the roof.

Note added in proof: It has been shown that C-terminal truncation of GroES after Asp 90 prevents heptamer formation⁵². □

Received 11 September; accepted 7 November 1995.

- Landry, S. J. & Gierasch, L. M. *Rev. Biophys. biomolec. Struct.* **23**, 645–669 (1994).
- Hartl, F. U., Hlodan, R. & Langer, T. *Trends biochem. Sci.* **19**, 20–25 (1994).
- Tilly, K. & Georgopoulos, C. *J. Bact.* **149**, 1082–1088 (1982).
- Lorimer, G. H. *Structure* **2**, 1125–1128 (1994).
- Ellis, R. J. *Curr. Biol.* **4**, 633–635 (1994).
- Braig, K. et al. *Nature* **371**, 578–586 (1994).
- Chen, S. et al. *Nature* **371**, 261–264 (1994).
- Langer, T., Pfeifer, G., Martin, J., Baumeister, W. & Hartl, F. U. *EMBO J.* **11**, 4757–4765 (1992).
- Fenton, W. A., Kashi, Y., Furtak, K. & Horwich, A. L. *Nature* **371**, 614–619 (1994).
- Schmidt, M. et al. *Science* **265**, 656–659 (1994).
- Martin, J., Mayhew, M., Langer, T. & Hartl, F. U. *Nature* **366**, 228–233 (1993).
- Todd, M. J., Viitanen, P. V. & Lorimer, G. H. *Science* **265**, 659–666 (1994).
- Landry, S. J., Zeilstra-Ryalls, J., Fayet, O., Georgopoulos, C. & Gierasch, L. M. *Nature* **364**, 255–258 (1993).
- Bochkareva, E. S., Lissin, N. M. & Girshovich, A. S. *Nature* **336**, 254–257 (1988).
- Gray, T. E. & Fersht, A. R. *J. molec. Biol.* **232**, 1197–1207 (1993).
- Schmidt, M., Buchner, J., Todd, M. J., Lorimer, G. H. & Viitanen, P. V. *J. biol. Chem.* **269**, 10304–10311 (1994).
- Martin, J. et al. *Nature* **352**, 36–42 (1991).
- Weissman, J. S. & Kim, P. S. *Science* **253**, 1386–1393 (1991).
- Jackson, G. S. et al. *Biochemistry* **32**, 2554–2563 (1993).
- Weissman, J. S., Kashi, Y., Fenton, W. A. & Horwich, A. L. *Cell* **78**, 693–702 (1994).
- Bochkareva, E. S. & Girshovich, A. S. *J. biol. Chem.* **267**, 25672–25675 (1992).
- Gray, T. E. & Fersht, A. R. *FEBS Lett.* **292**, 254–258 (1991).
- Bochkareva, E. S., Lissin, N. M., Flynn, G. C., Rothman, J. E. & Girshovich, A. S. *J. biol. Chem.* **267**, 6796–6800 (1992).
- Vellieux, F. M. D., Hunt, J. F., Roy, S. & Read, R. J. *J. appl. Crystallogr.* **28**, 347–351 (1995).
- Fitzgerald, P. M. D. *J. appl. Crystallogr.* **21**, 273–278 (1988).
- Rossmann, M. G. & Blow, D. M. *Acta crystallogr.* **15**, 24–31 (1962).
- Sander, C. & Schneider, R. *Proteins* **9**, 56–68 (1991).
- Rost, B. & Sander, C. *Proteins* **19**, 55–72 (1994).
- Wilmot, C. M. & Thornton, J. M. *Protein Engng* **3**, 479–493 (1990).
- Nicholls, A. *GRASP: Graphical Representation and Analysis Surface Properties* (Columbia University, New York, 1992).
- Connolly, M. L. *J. appl. Crystallogr.* **16**, 548–558 (1983).
- Collaborative Computational Project, Number 4 *Acta Crystallogr. D* **50**, 760–763 (1994).
- Herzberg, O. & Mout, J. *Proteins* **11**, 223–229 (1991).
- Janin, J., Miller, S. & Chothia, C. *J. molec. Biol.* **204**, 155–164 (1988).
- Zondlo, J., Fisher, K. E., Lin, Z., Ducote, K. R. & Eisenstein, E. *Biochemistry* **34**, 10334–10339 (1995).
- Lawrence, M. C. & Colman, P. M. *J. molec. Biol.* **234**, 946–950 (1993).
- Jencks, W. P. *Adv. Enzymol.* **43**, 219–410 (1975).
- Drenth, J. *Principles of Protein X-ray Crystallography* (Springer, New York, 1994).
- Steigemann, W. *Int. Union Crystallogr. crystallogr. Symp.* **5**, 115–125 (1991).
- Otwinski, Z. in *Isomorphous Replacement and Anomalous Scattering* (eds Wolf, W., Evans, P. R. & Leslie, A. G. W.) 80–86 (SERC Daresbury Laboratory, Warrington, 1991).
- Brunger, A. T. *X-PLOR, Version 3.1: A System for Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992).
- Kabsch, W. *J. appl. Crystallogr.* **21**, 916–924 (1988).
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* **A47**, 110–119 (1991).
- Engh, R. A. & Huber, R. *Acta crystallogr.* **A47**, 392–400 (1991).
- Wang, B. C. *Meth. Enzym.* **115**, 90–112 (1985).
- Leslie, A. G. W. *Acta crystallogr.* **A43**, 134–136 (1986).
- Kleywegt, G. J. & Jones, T. A. *Acta crystallogr.* **D50**, 178–185 (1994).
- Kabsch, W. & Sander, C. *Biopolymers* **22**, 2577–2637 (1983).
- Kraulis, P. J. *J. appl. Crystallogr.* **24**, 946–950 (1991).
- Merritt, E. A. & Murphy, M. E. P. *Acta crystallogr.* **D50**, 869–873 (1994).
- Weissman, J. S. et al. *Cell* **83**, 577–587 (1995).
- Seale, J. W. & Horowitz, P. M. *J. biol. Chem.* **270**, 30268–30270 (1995).

ACKNOWLEDGEMENTS. We thank T. Wingert for technical assistance with protein purification; H. Saibil for Fig. 1e; A. J. Hudspeth for a critical review of the manuscript; O. Ortiz for producing the dye-sublimation prints; and the members of the U.T. Southwestern crystallography group for advice and assistance. J.F.H. was supported by a postdoctoral fellowship from the Jane Coffin Childs Memorial Fund for Medical Research. S.J.L. acknowledges support from LEQSF and NSF.

Magnetic spiral arms in the galaxy NGC6946

R. Beck & P. Hoernes

Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany

MAGNETIC fields pervade spiral galaxies¹—at least all those searched so far; these fields have a substantial energy density, and hence could play an important role in galaxy evolution. Dynamo theory has been used for many years to explain the presence and overall structure of galactic magnetic fields, through the amplification of a weak seed field^{2,3}. Here we report the observation of two ‘magnetic spiral arms’ in the nearby galaxy NGC6946, lying between the optical spiral arms. This is surprising because dynamo action is thought to be related to star formation activity⁴, which is concentrated within or in the leading edges of the optical spiral arms. The magnetic spiral arms are about 500–1,000 parsecs wide and more than 12 kiloparsecs long, and have greater symmetry than the optical arms. This organized structure probably reflects the signature of some global mechanism relating to magnetic field generation, but no current theory—in particular dynamo theory in its present form—is able to explain this phenomenon.

NGC6946 is one of the apparently largest spiral galaxies in the sky owing to its proximity (distance determinations range between 5 and 10 Mpc; we assume 7 Mpc throughout this Letter). Its spiral structure (seen almost face-on with an inclination angle of $\sim 30^\circ$ to the sky plane) is multiple-armed and even irregular in some regions. Lacking a central stellar bar and a companion galaxy, the density waves are only of modest strength. Radio polarization observations with the Effelsberg 100-m telescope revealed an overall spiral pattern of the interstellar magnetic field in NGC6946 (refs 5–7), but the angular resolution was insufficient to study the effects of spiral arms on the magnetic field. Observations with the Very Large Array (VLA) at 20-cm wavelength with 42-arcsec resolution⁸ were hampered by Faraday depolarization caused by strong turbulent magnetic fields ($\sim 12 \mu\text{G}$ on average, up to $20 \mu\text{G}$ in the optical spiral arms) and the ionized interstellar gas in the disk. Excess Faraday rotation and depolarization at a wavelength of 20 cm in the southwestern quadrant give an indication of some open field lines bending into the halo, a ‘galactic coronal hole’, which obtains further support from a minimum in the coronal X-ray emission as observed by the Rosat satellite⁹.

We observed NGC6946 with the VLA of the National Radio Astronomy Observatory in its D configuration for 9.5 hours at 6.2-cm wavelength. At this wavelength, synchrotron emission dominates and the effects of Faraday rotation and Faraday depolarization on the polarization vectors are small, so that we can directly study the properties of the interstellar magnetic fields. The telescope data were calibrated, cleaned and self-calibrated with the standard routines of the AIPS software package. The synthesized beam width of 12.5 arcsec gives a resolution of 420 pc at the assumed distance of NGC6946.

As an interferometer the VLA is not sensitive to large-scale spatial structures: in our original maps, all emission on scales larger than ~ 7 arcmin was

invisible. However, we were able to add the missing interferometer spacings to the maps in total power and linear polarization with the help of Effelsberg observations at almost the same wavelength⁶. This was achieved by combining the clean VLA map and the Effelsberg map in the Fourier plane of spatial frequencies using IMERG, a task in the Software Development Environment (SDE) of the NRAO¹⁰. IMERG deconvolves both maps from their gaussian-shaped beams, performs a weighted addition of the Fourier-transformed maps, convolves with the gaussian VLA beam, and transforms the resulting map back into the image plane. The weights for the addition were chosen in such a way that the large-scale emission corresponding to small spacings in the Fourier plane were taken from the Effelsberg map, while the small-scale emission corresponding to large spacings came from the VLA map. In the domain of intermediate spacings seen by both telescopes a linear interpolation was used. Owing to the different responses of the two telescopes to point-like sources, IMERG introduces ring-like distortions around point-like sources which limits the dynamic range to $\sim 2,000$ times the r.m.s. noise¹¹. Therefore, the bright central source and another bright point source in the west were subtracted before the combination, and added again afterwards.

The VLA images alone contain only $\sim 27\%$ of the total flux density and $\sim 40\%$ of the linearly polarized flux density visible in the combined maps. Hence only the combination of single-dish and synthesis data reveals the full extent of the radio emission.

The total synchrotron radio emission (Fig. 1) is rather diffuse, probably emerging from a thick disk, plus a component which follows the multiple-armed distribution of the stars and the interstellar gas. In contrast, the polarized radio emission (Fig. 2)

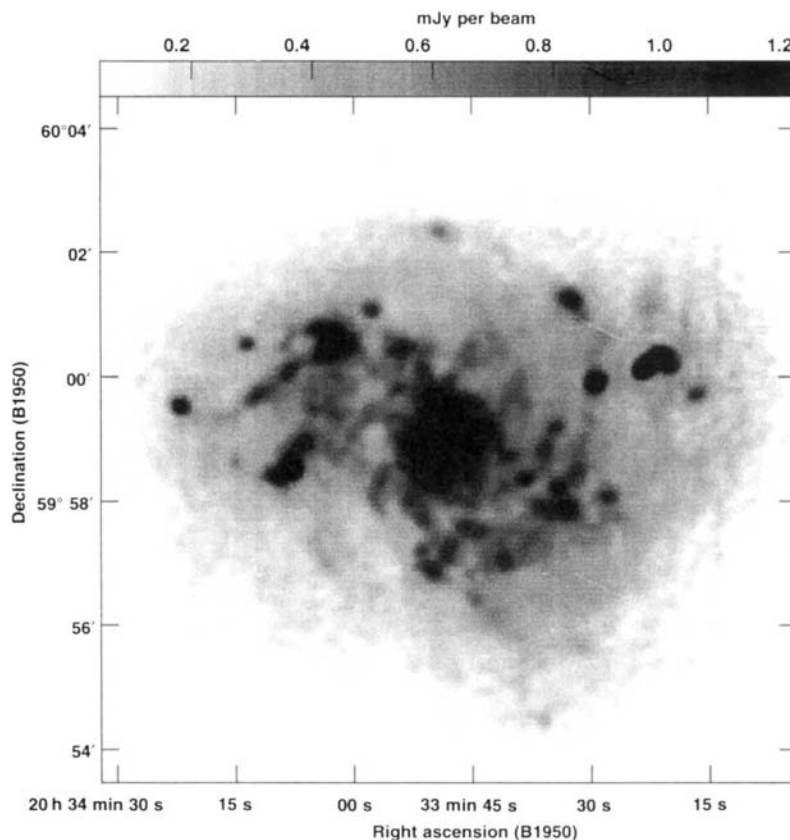
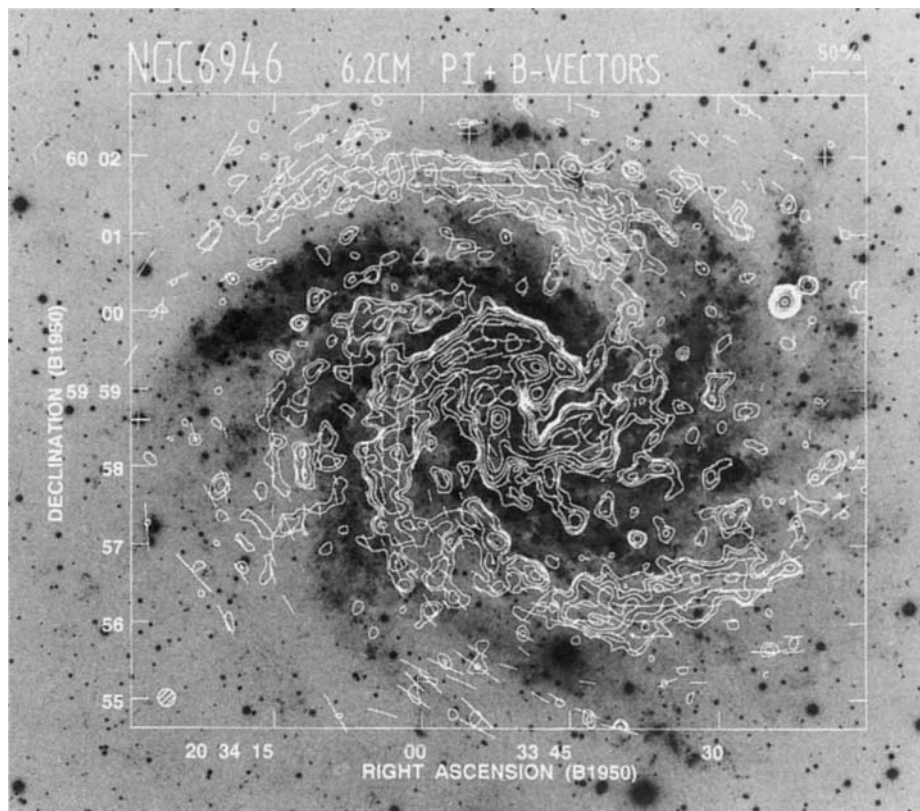


FIG. 1 Total radio emission of NGC6946 at 6.2-cm wavelength. The grey-scale (top of image) covers the range 0.05–1.2 mJy per beam. The r.m.s. noise is $\sim 20 \mu\text{Jy}$ per beam.

FIG. 2 Linearly polarized radio emission (PI) of NGC6946 at 6.2-cm wavelength, overlaid onto an optical image. Contour levels are $(30 \mu\text{Jy per beam}) \times 2^{n/2}$ ($n = 0, 1, 2, \dots$). The r.m.s. noise is $\sim 10 \mu\text{Jy per beam}$. The lengths of the B vectors (short white lines) are proportional to the degree of polarization.



forms a surprisingly symmetric spiral pattern with two main arms containing highly aligned magnetic fields, outside (but perfectly parallel to) the optical spiral arms. Another pair of rudimentary magnetic arms is visible in the eastern and western parts of the galaxy. The spiral orientation of the field in the main magnetic arms can be followed from the centre to the outskirts of the galaxy. Southwest of the nucleus, the two main magnetic arms get so close that their (perpendicular) vectors cause beam depolarization along a trough.

The typical width of the two main magnetic spiral arms is only ~ 500 – $1,000$ pc (after correction for the limited resolution) compared with their length of at least 12 kpc each. The degree of polarization is typically 30%. After subtraction of the diffuse, mostly unpolarized background (see Fig. 1), the degree of polarization reaches 65% in the northern magnetic arm (Fig. 3) and 55% in the southern one. Compared with the theoretical limit of 74% polarization for synchrotron emission (assuming an isotropic distribution of cosmic-ray electrons), we conclude that the field in the magnetic arms is almost perfectly aligned. The strength of this regular field is ~ 3 – $13 \mu\text{G}$ (assuming energy equipartition between magnetic fields and cosmic rays, and a ratio of 100 between the energy densities of cosmic-ray protons and electrons).

The magnetic spiral arms are situated between two optical spiral arms which are also seen in the total radio emission (Fig. 1). As a representative example, Fig. 3 shows the distribution along a line cutting through the northern magnetic arm, avoiding discrete sources. The central position of the magnetic arm is not exactly in the middle between the total-power arms, but shifted towards the outer arm. The magnetic arms are also visible at a wavelength of 20 cm (ref. 8), although much less prominent owing to the lower angular resolution and Faraday depolarization. Although the density of warm gas is low in the interarm regions, Faraday effects at 20 cm reveal the occurrence of ionized gas also in the interarm regions, possibly the hot gas observed in X-rays⁹.

Peaks of polarized emission in interarm regions are a general phenomenon in spiral galaxies¹. This has been interpreted as the absence of field tangling and small Faraday depolarization. Field

tangling occurs only in the optical spiral arms owing to supernova explosions and turbulent motions of gas clouds⁸. However, our new observations clearly show that the polarized features are significantly narrower than the interarm space, especially in the outer regions of the galaxy where the separation of the optical arms increases. Furthermore, the magnetic arms are visible as separate peaks in total-power radio emission (see Fig. 3). Hence the magnetic arms are an independent phenomenon, and not only due to a different degree of field tangling or Faraday depolarization inside and outside the optical spiral arms.

IC342 is a gas-rich galaxy similar to NGC6946 and also shows magnetic arms¹², less pronounced and less regular than in NGC6946, but extending to an even larger distance from the galaxy's centre¹³. The same is true for M83¹⁴. In M31, M51 and M81, however, no indications for magnetic arms have been found¹. Thus it seems possible that the phenomenon of magnetic spiral arms is common among late-type, gas-rich, non-interacting galaxies, but has not been noticed so far because of the limited spatial resolution of the previous observations.

The degree of symmetry of the regular field in NGC6946 is much higher than that of the gas and stars. Such a field pattern must be generated by a global mechanism steadily operating over times longer than a rotation period. Given our current understanding of global structure in spiral galaxies, there seems to be no adequate theory to account for our results. The following four alternatives, acting on a global basis, do not seem to work.

Density waves. An almost perfect field alignment parallel to the spiral arms needs a strong shock in order to deflect the field components perpendicular to the shock. Compression zones due to such a supersonic density-wave flow generally occur at the inner edges of spiral arms, and are marked by dust lanes and strong peaks in total radio continuum. The magnetic arms in NGC6946, however, are not located at the inner arm edges and not traced by dust lanes or bright ridges in the total-power map.

Instabilities. Magnetic arms could be the result of thermal instabilities¹⁵. But the typical width of such an instability is ~ 1 pc for a $10\text{-}\mu\text{G}$ magnetic field, much too small to explain the

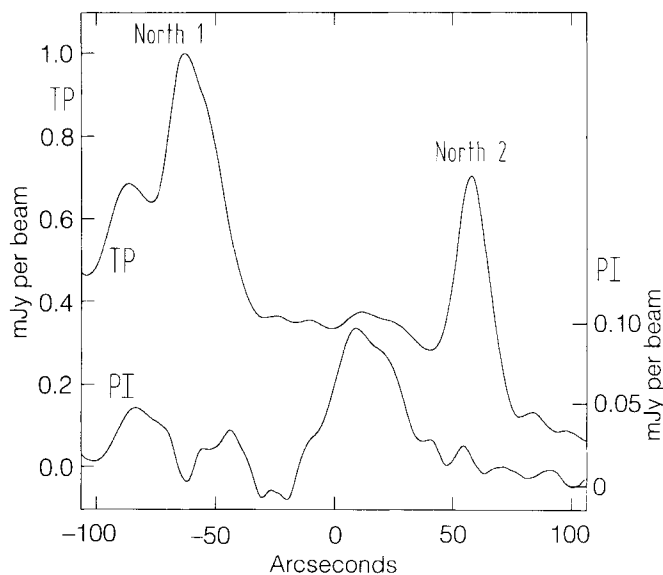


FIG. 3 Cut through the total-power (TP; upper curve, left-hand vertical axis) and polarized radio emission (PI; lower curve, right-hand vertical axis) of the northern spiral arms of NGC6946, centred on right ascension 20 h 33 min 51.87 s, declination 60° 01' 25" (B1950.0) and with a position angle of -21.5° . 'North 1' and 'North 2' indicate the positions of the optical/total-power spiral arms.

present observations. Magnetic buoyancy instabilities¹⁶ are similarly improbable because there are no hints that the magnetic arms in NGC6946 bend above the plane.

Magnetic reconnection. The interarm space may be the location of a current sheet where the field direction reverses. Current sheets are the location of electron acceleration¹⁷ which in turn may emit synchrotron emission. However, with $\sim 10\text{-}\mu\text{G}$ interarm fields in NGC6946, particles can hardly reach high enough

Lorentz factors (see equation (29) in ref. 18). Furthermore, the large-scale Faraday rotation pattern is not in favour of field reversals⁷.

Dynamo theory. A global field pattern requires a global generation mechanism such as the galactic dynamo^{2,3}. Most models assume axial symmetry in velocity and distribution of the ionized gas, and predict a dominant axisymmetric ($m = 0$) magnetic mode. The $m = 1$ dynamo mode has a bisymmetric two-armed spiral structure^{2,3}, but of much larger width than the magnetic arms in NGC6946. The inclusion of non-axisymmetric effects (like spiral arms) in gas density, velocity field or other dynamo parameters may generate a superposition of even modes ($m = 0, m = 2, \dots$)^{2,19} which may lead to two dominant magnetic spiral arms. However, the observed location of such arms in regions of minimum star-formation rate still awaits explanation. $\square$

Received 19 July; accepted 16 November 1995.

1. Beck, R. in *The Cosmic Dynamo* (eds Krause, F., Rädler, K. H. & Rüdiger, G.) 283–297 (IAU Symp. No. 157, Kluwer Academic, Dordrecht, 1993).
2. Ruzmaikin, A., Sokoloff, D. & Shukurov, A. *Nature* **336**, 341–347 (1988).
3. Elstner, D., Meinel, R. & Beck, R. *Astr. Astrophys. Suppl. Ser.* **94**, 587–600 (1992).
4. Ko, C. M. & Parker, E. N. *Astrophys. J.* **341**, 828–831 (1989).
5. Klein, U., Beck, R., Buczkowski, U. R. & Wiebeleski, R. *Astr. Astrophys.* **108**, 176–187 (1982).
6. Harnett, J. I., Beck, R. & Buczkowski, U. *Astr. Astrophys.* **208**, 32–38 (1989).
7. Ehle, M. & Beck, R. *Astr. Astrophys.* **273**, 45–64 (1993).
8. Beck, R. *Astr. Astrophys.* **251**, 15–26 (1991).
9. Schlegel, E. M. *Astrophys. J.* **434**, 523–535 (1994).
10. Cornwell, T. J., Briggs, D. S. & Holdaway, M. A. *User's Guide to SDE* (NRAO, Socorro, 1995).
11. Holdaway, M. A. *Mosaicing with Very High Dynamic Range* (Millimeter Array Memo No. 73, NRAO, Socorro, 1992).
12. Krause, M. in *The Cosmic Dynamo* (eds Krause, F., Rädler, K. H. & Rüdiger, G.) 305–310 (IAU Symp. No. 157, Kluwer Academic, Dordrecht, 1993).
13. Krause, M., Hummel, E. & Beck, R. *Astr. Astrophys.* **217**, 4–16 (1989).
14. Sukumar, S. & Allen, R. J. *Nature* **340**, 537–539 (1989).
15. Rosso, F. & Pelletier, G. *Astr. Astrophys.* **270**, 416–425 (1993).
16. Kleorin, N. I. & Rogachevskii, I. V. in *Plasma Astrophysics 21–23* (ESA SP-311, European Space Agency, Paris, 1990).
17. Lesch, H. & Reich, W. *Astr. Astrophys.* **264**, 493–499 (1992).
18. Lesch, H. & Pohl, M. *Astr. Astrophys.* **254**, 29–38 (1992).
19. Moss, D., Brandenburg, A., Donner, K. J. & Thomasson, M. *Astrophys. J.* **409**, 179–189 (1993).

ACKNOWLEDGEMENTS. We thank M. Holdaway and D. Briggs for their fast and reliable support on SDE-related problems. We are grateful to the NRAO for observing time at the VLA. The NRAO is operated by Associated Universities Inc., under cooperative agreement with the US NSF.

Avalanche dynamics in a pile of rice

Vidar Frette*, Kim Christensen,
Anders Malthe-Sørensen, Jens Feder,
Torstein Jøssang & Paul Meakin

Department of Physics, University of Oslo, PO Box 1048, Blindern, N-0316 Oslo, Norway

THE idea of self-organized criticality¹ (SOC) is commonly illustrated conceptually with avalanches in a pile of sand grains. The grains are dropped onto a pile one by one, and the pile ultimately reaches a stationary 'critical' state in which its slope fluctuates about a constant angle of repose, with each new grain being capable of inducing an avalanche on any of the relevant size scales. Some numerical models of sand-pile dynamics do show SOC^{1–8}, but the behaviour of real sand piles remains ambiguous^{9–18}. Here we describe experiments on a granular system—a pile of rice—in which the dynamics exhibit self-organized critical behaviour in one case (for grains with a large aspect ratio) but not in another (for less elongated grains). These results show that SOC is not as 'universal' and insensitive to the details of a system

as was initially supposed¹, but that instead its occurrence depends on the detailed mechanism of energy dissipation.

We have studied the dynamics of piles of rice for several system sizes (Fig. 1). Grains of rice were slowly fed into the gap between two vertical, parallel plates. Mass accumulated in the pile until it reached a quasi-stationary state. Under continued addition, avalanches redistributed mass along the surface layer of the heap and transported mass out of the system, thereby changing the profile. The anisotropy of the grains (see Table 1) gave rise to a variety of packing configurations⁷ (Fig. 1). The anisotropy restricted the way the grains moved down the slope, enhanced the frictional contact and, to a large extent, suppressed inertia effects. The size of an avalanche was defined as the energy dissipated between two consecutive profiles (Fig. 2a,b). A temporal sequence of avalanche sizes for rice A is shown in Fig. 2c. There was a large variation in the avalanche sizes and an apparently chaotic structure in time. We concentrate here on the statistics of the avalanche sizes.

Based on experience with a wide range of complex systems, it is reasonable to expect that the probability density $P(E, L)$, where $P(E, L)dE$ is the probability that an avalanche with energy dissipation between E and $E + dE$ will occur in a system of size L , will have the form

$$P(E, L) = L^{-\beta} f(E/L^{\nu}) \quad (1)$$

where the scaling function f and the scaling exponents β and ν are to be determined. From the normalization $\int_0^\infty P(E, L)dE = 1$ it follows that $\beta = \nu$. Further, in the stationary state, the average

* Present address: Department of Physics of Complex Systems, The Weizmann Institute of Science, Rehovot 76100, Israel.

TABLE 1 Types of rice used in experiments

	Rice A	Rice B	Rice C
Name	Geisha* <i>naturris</i>	Geisha* <i>grøtris</i>	Geisha* <i>middagsris</i>
Description	Unpolished (rough)	Polished (smooth)	Polished (smooth)
Length, δ	7.6 ± 0.9 mm	4.8 ± 0.4 mm	6.9 ± 1.0 mm
Width	2.0 ± 0.1 mm	2.4 ± 0.2 mm	1.9 ± 0.1 mm
Aspect ratio	3.8	2	3.6
Average mass, m	0.02066 g	0.01840 g	0.01874 g
No. density	37 ± 1 grains cm^{-3}	43 ± 1 grains cm^{-3}	44 ± 1 grains cm^{-3}
Energy unit, ε	1.54 μJ	0.867 μJ	1.27 μJ
Angle of repose	$48.4^\circ \pm 1.2^\circ$	$46.6^\circ \pm 0.8^\circ$	$51.1^\circ \pm 2.0^\circ$

The large variation in length for rice A and C adds to the inhomogeneity of the system (Fig. 1). The average grain mass, m , was obtained by weighing 2,500 grains. The number density is the average number of grains per unit volume. The energy unit is $\varepsilon = mg\delta$, where g is the acceleration of gravity and δ is the grain length. Throughout the Letter we express the energy dissipation in units of ε , that is, by the dimensionless quantity $E = \mathcal{E}/\varepsilon$, where $\mathcal{E}$ is the energy dissipation of an avalanche. Similarly, we express the system size $\mathcal{L}$ by the dimensionless number $L = \mathcal{L}/\delta$. The average pile height at the vertical wall was used to estimate the angle of repose. The standard deviation in this angle reflects the observed height fluctuations.

* Geisha is a registered trademark from Stabburet a.s. (Trollåsen, Norway).

dissipated energy $\langle E \rangle = \int_0^\infty EP(E,L)dE / \int_0^\infty P(E,L)dE \propto L^\nu$ is equal to the average added potential energy. Because the angle of repose varied only weakly with system size, the added potential energy was proportional to L . Thus $\langle E \rangle \propto L$, and $\nu = 1$.

Figure 3a shows the probability density $P(E, L)$ for rice A with system sizes of $L = 16, 33, 66$ and 105 (see Table 1). The behaviour can be described by a scaling function of the form $f(x) = \text{const. for } x < 1$, and $f(x) \propto x^{-2}$ for $x > 1$. Figure 3b shows a finite-size scaling plot of these results with $\beta = \nu = 1$. There is a reasonably good data collapse. The exponent α was estimated from fits in the range $6 \leq E/L \leq 300$. We found size-dependent corrections to scaling for the exponent α of the form $\alpha = \alpha_\infty + l/L$, with $l \approx 8.6$, where $\alpha_\infty \approx 2.06$ is the power-law exponent in a system with $L \rightarrow \infty$. From data covering the range $10 \leq E/L \leq 100$, we found $\alpha_\infty \approx 2.02$ and $l \approx 10.4$. We conclude that α_∞ is slightly above 2 and $l = 9.5 \pm 1$. The avalanche

dynamics in the rice pile with the elongated rice A is consistent with a self-organized critical process.

Figure 4a shows the probability density $P(E, L)$ for rice B with $L = 26, 52$ and 104 . Using $\nu = \beta = 1$, we find that the probability densities are consistent with a stretched-exponential scaling function, $f(x) \propto \exp(-(x/x^*)^\gamma)$ with $\gamma \approx 0.43$ and $x^* \approx 0.45$ (Fig. 4b). A characteristic avalanche size $E^* = x^*L$ for the dynamics of rice B appears, which is inconsistent with the idea of SOC.

We conclude that SOC behaviour is not a universal characteristic of slowly driven granular media. A difference in the detailed relaxation mechanisms for rice A and the more isotropic rice B might explain the crossover from a critical to a non-critical behaviour. For the elongated rice A, the relaxation dynamics was dominated by (1) ‘domains’ of grains that moved slowly, coherently and intermittently, while keeping most of their ‘solid-like’ internal structure, and (2) many ‘individual’ grains that

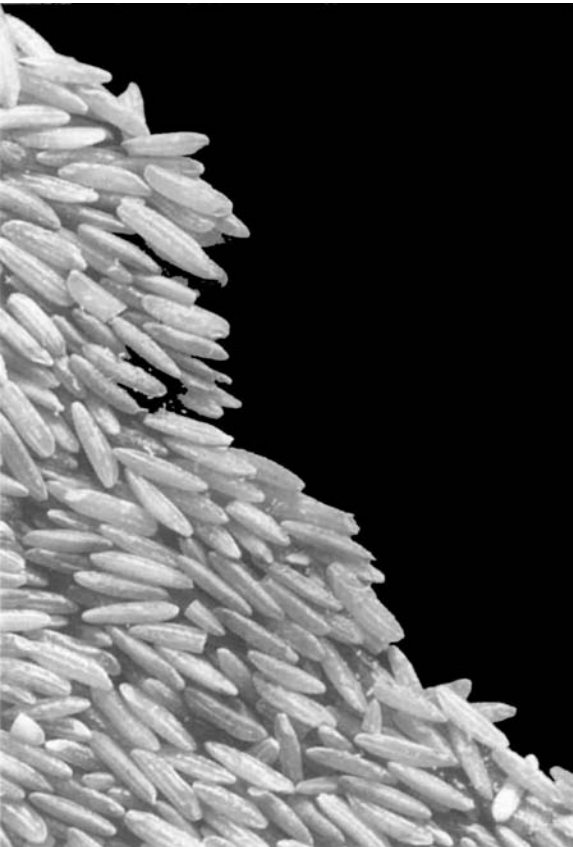


FIG. 1 A close-up photograph of a pile of rice A (see Table 1). Piles of rice were confined between two 5-mm-thick glass plates supported by 100×120 cm polymethylmethacrylate plates (15 mm thick). Aluminium rods formed the vertical wall and the variable one-dimensional base of the pile (Fig. 2a, b). All the experiments were performed with a constant plate-separation to grain-length ratio of ~ 0.8 . In an extensive series of experiments, we did not find any qualitative changes as the plate separation was varied. As shown, most of the grains were aligned along the flow direction, but there was a significant spread in the angular orientation of the grains, and a variety of packing configurations were obtained. As a result, the local slope varied considerably along the profile. These ‘patterns’ of steep and flat portions in the profile changed substantially in time. The dynamics of the rice piles was recorded with a photometric CCD (charge-coupled device) camera (CH250 from Photometrics, Tucson, Arizona) with a spatial resolution of $2,000 \times 2,000$ pixels, and 12-bit grey-level resolution. For each system size, we chose lens and camera distance so that a $2,000 \times 500$ pixels picture segment covered the active zone of the pile. Grains were injected at the vertical wall at a rate of $20 \text{ grains min}^{-1}$. Frames were taken at 15 s intervals and the profiles (typically $\sim 2,800$ pixels) were identified. Each experiment lasted ~ 42 h and consisted of 10,000 profiles. Our choice of feeding and sampling rates represent a compromise between the ideal of infinitely slow driving and the need to explore much of the large configuration space of the system.

flowed like a fluid along the profile. Occasionally, almost stagnant domains suddenly collapsed and released a flow, but frequently, the avalanches were initiated from above by flowing grains or from below through undermining as regions further down slid off. Series of such underminings constituted back-avalanches² that slowly and intermittently propagated upwards on the pile. Sometimes, single grains bounced through the system, but the frequency of these bouncing events decreased when L increased. This may have contributed to the decrease of the power-law exponent α as the system size was increased.

Whereas grains of rice A slide along the surface, the more spherical rice B tended to roll. The dynamics of rice B consisted mainly of single grains that bounced rapidly down the slope or many 'individual' grains that flowed like a fluid along the profile. The grains often accumulated kinetic energy as they moved down the slope and were not easily halted. Typically, they propagated halfway through the system, independent of system size. Thus the rolling grains of rice B frequently tested the stability throughout the slope of the pile, and the fluctuations around the 'mean' profile were smaller than for the elongated

rice A. Furthermore, this 'inertia effect' led to a non-local process whereas in the numerical models displaying SOC and the rice pile with elongated grains, the dynamics was dominated by local mechanisms.

In an experiment with a third rice C (elongated like rice A, smooth like rice B; Table 1) we observed sliding mechanisms and profile shapes similar to rice A and the probability density was again a power law. This suggests that the statistics of the relaxation events is dominated by the shape rather than the surface properties of the grains. A large aspect ratio leads to a 'rough' profile and a sliding grain motion with a higher effective friction, which is crucial to getting 'critical dynamics'. It would be interesting to find a computer model in which a crossover from critical to non-critical behaviour is observed as a function of a relevant parameter, such as the ratio between the dissipation and the gain in kinetic energy as the grains move down the slope.

A reanalysis¹⁸ of all the previous experiments⁹⁻¹⁷ on granular systems shows that the flow over the rim of the pile ('drop numbers') has a stretched-exponential distribution inconsistent with the SOC hypothesis. By measuring the internal avalanches

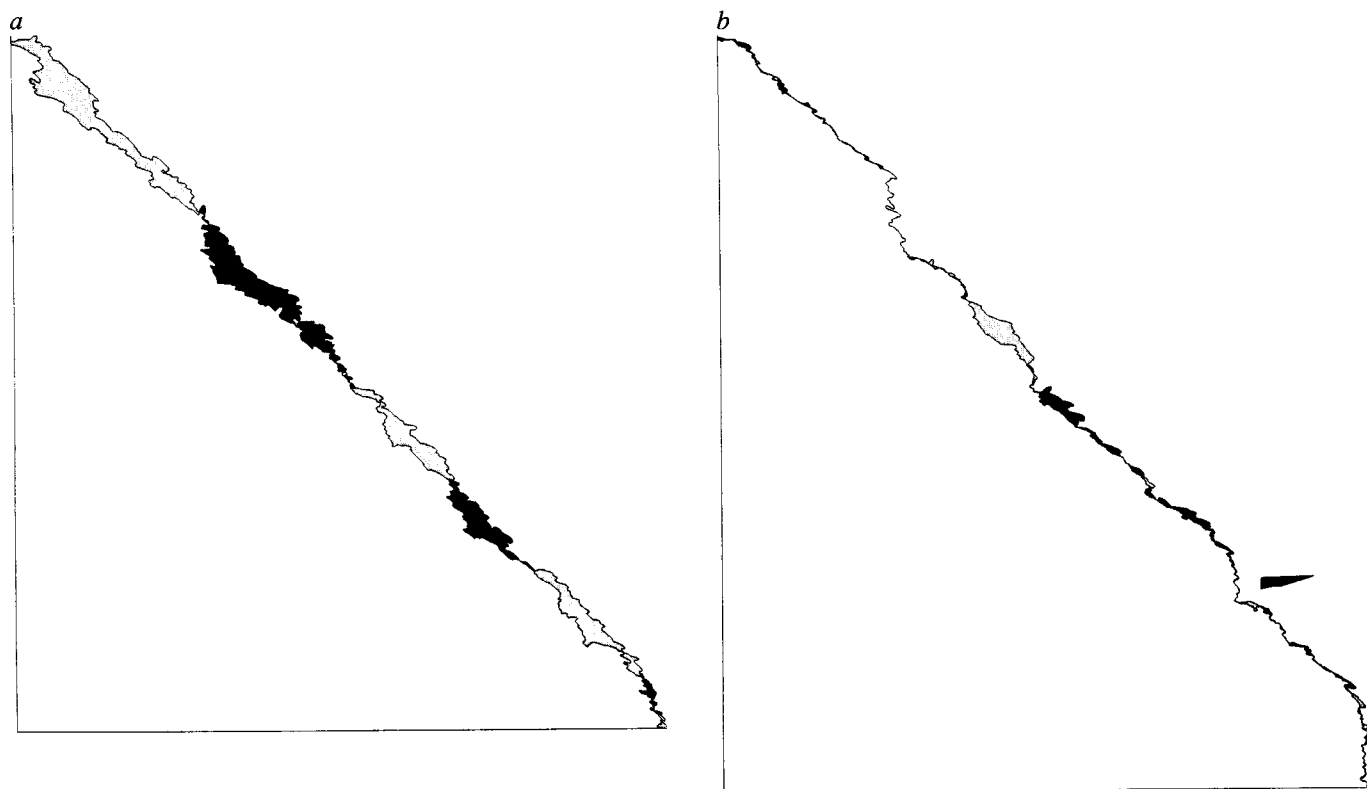


FIG. 2 *a, b*, Two avalanches of different sizes from an experiment with rice A at a plate separation of $d/\delta = 0.79$ and system size of $L = 33$ expressed in terms of the grain length δ (see Table 1) (the baseline length in *a* and *b* was 25 cm = 1,548 pixels). In each case, two successive profiles have been plotted together. The areas enclosed by the two profiles represent the reorganization of mass in the pile. Loss (gain) of mass is indicated by grey (black) areas. Whenever grains have moved down the slope, potential energy has been lost and we measured the redistribution of mass in terms of dissipated potential energy. The sizes of these avalanches are 2,868 (*a*) and 294 (*b*). The energy dissipation is expressed in terms of $\varepsilon = mg\delta = 1.54 \mu\text{J}$ (see Table 1). Most of the grains slid some distance inside the pile, and only a smaller fraction reached the rim and left the system ('drop event'). The internal reorganizations (avalanches) reflect the response to the driving in a more direct way than the drop events. *c*, Temporal sequence of avalanche sizes for the experiment specified above. The size of each avalanche, rescaled by the maximal avalanche $E_{\max}$ for the entire experiment, is displayed as a vertical line. The sequence shown is a 2.5-h segment consisting of 600 avalanches. The arrows indicate the events shown in *a* and *b*.

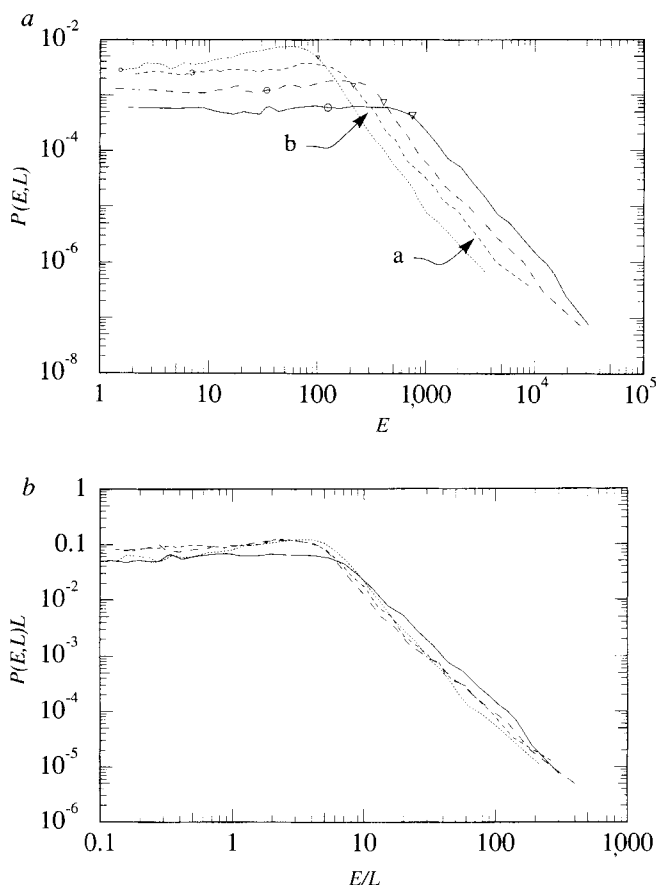


FIG. 3 a, Log-log plot of the probability density $P(E, L)$ as a function of energy dissipation E for rice A, a plate separation of $d/\delta = 0.79$, and system sizes $L = 16, 33, 66$ and 105 , marked with lines of increasing dash length. The energy dissipation is expressed in terms of $\varepsilon = 1.54 \mu\text{J}$ and the system size in terms of the grain length δ (see Table 1). The $L = 16$ experiment was performed twice with very similar results for the probability density. The probability densities are constant for 'small' values of E and have a power-law form $P(E, L) \propto E^{-\alpha}$, extending over $\sim 1\frac{1}{2}$ decades, for 'large' E , thus there is no characteristic avalanche size. The arrows show the avalanche sizes corresponding to the reorganization events in Fig. 2a, b. The average added energy between two profiles is indicated by triangles. This is taken into account when calculating the avalanche size. The noise level was determined by analysing a series of pictures (typically 5,000) of static profiles (no grain injection). The resulting probability densities have approximate exponential form with average 'avalanche' sizes indicated by circles. We were not able to get reliable results with our equipment beyond $L = 105$ owing to limited resolution. In an $L = 125$ experiment (not shown), the finer details such as the gaps between single grains at the surface of the pile were lost and the noise extended into the power-law range. However, the qualitative behaviour of the rice piles did not change as the system size was increased beyond $L = 105$. We measured the drop number for system sizes up to $L = 263$ and found no indication of a crossover to the oscillatory behaviour observed in previous experiments on sandpiles^{9,10}. b, A finite-size scaling plot of the data in a using equation (1) with $\beta = \nu = 1$. The scaling function has the form $f(x) = \text{const.}$ for $x < 1$ and $f(x) \propto x^{-\alpha}$ for $x > 1$, with $\alpha \approx 2.04$. There are no obvious cutoffs in the scaling function, and the power-law behaviour extends up to the largest avalanche sizes measured in each case. In the scaling argument, a cutoff is unnecessary when the power-law exponent $\alpha > 2$.

for systems of different sizes, we have found that the theory of self-organized criticality applies to some slowly driven granular systems but is not a universal phenomenon. $\square$

Received 12 January; accepted 13 November 1995.

1. Bak, P., Tang, C. & Wiesenfeld, K. *Phys. Rev. Lett.* **59**, 381–384 (1987); *Phys. Rev.* **A38**, 364–374 (1988).

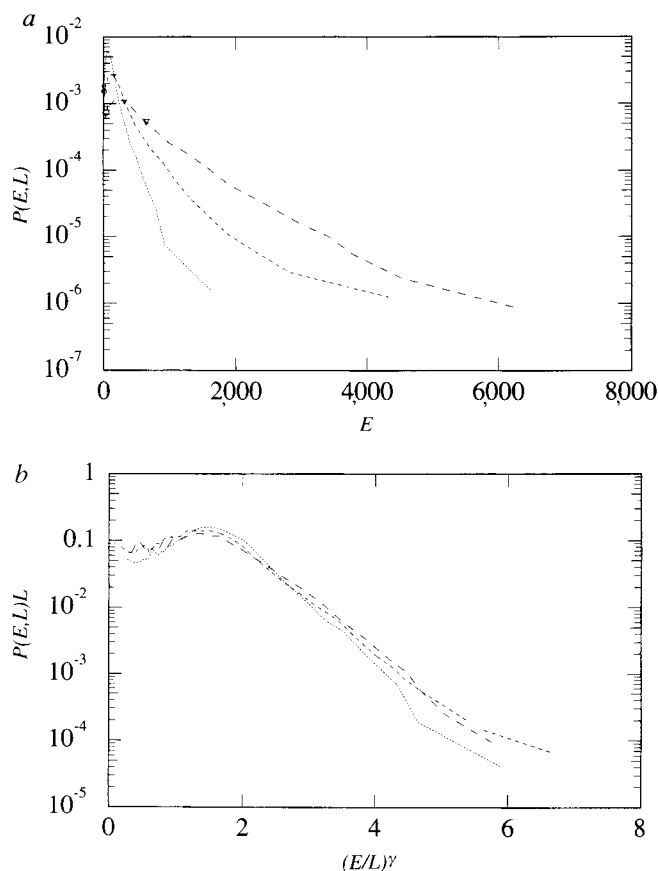


FIG. 4 a, Semi-log plot of the probability density $P(E, L)$ as a function of energy dissipation E for rice B with a plate separation of $d/\delta = 0.83$, and system sizes $L = 26, 52$ and 104 , indicated by increasing dash length for the curves. The energy was measured in terms of $\varepsilon = 0.867 \mu\text{J}$ and the system size in terms of the grain length δ (see Table 1). The average added energy between two profiles is indicated by triangles. This is taken into account when calculating the energy dissipation. The noise levels, determined in the manner explained in Fig. 3 legend, are indicated by circles. b, Scaling plot of the data in a with $\beta = \nu = 1$. The form of the scaling function in equation (1) is consistent with a stretched exponential, $f(x) \propto \exp(-(x/x^*)^\gamma)$ with $\gamma \approx 0.43 \pm 0.03$ and $x^* \approx 0.45 \pm 0.09$.

2. Kadanoff, L. P., Nagel, S. R., Wu, L. & Zhou, S.-m. *Phys. Rev.* **A39**, 6524–6537 (1989).
3. Carlson, J. M. & Langer, J. S. *Phys. Rev. Lett.* **62**, 2632–2635 (1989).
4. Feder, H. J. S. & Feder, J. *Phys. Rev. Lett.* **66**, 2669–2672 (1991); *Phys. Rev. Lett.* **67**, 283(E) (1991).
5. Olami, Z., Feder, H. J. S. & Christensen, K. *Phys. Rev. Lett.* **68**, 1244–1247 (1992).
6. Drossel, B. & Schwabl, F. *Phys. Rev. Lett.* **69**, 1629–1632 (1992).
7. Frette, V. *Phys. Rev. Lett.* **70**, 2762–2765 (1993).
8. Bak, P. & Sneppen, K. *Phys. Rev. Lett.* **71**, 4083–4086 (1993).
9. Jaeger, H. M., Liu, C.-h. & Nagel, S. R. *Phys. Rev. Lett.* **62**, 40–43 (1989).
10. Held, G. A. et al. *Phys. Rev. Lett.* **65**, 1120–1123 (1990).
11. Evesque, P. *Phys. Rev.* **A43**, 2720–2740 (1991).
12. Evesque, P., Fargue, D., Habib, P., Luong, M. P. & Porion, P. *Phys. Rev.* **E47**, 2326–2332 (1993).
13. Bretz, M., Cunningham, J. B., Kurczynski, P. L. & Nori, F. *Phys. Rev. Lett.* **69**, 2431–2434 (1992).
14. Morales-Gamboa, E., Lomnitz-Adler, J., Romero-Rochin, V., Chicharro-Serra, R. & Peralta-Fabi, R. *Phys. Rev.* **E47**, R2229–R2232 (1993).
15. Grumbacher, S. K., McEwen, K. M., Halverson, D. A., Jacobs, D. T. & Lindner, J. *Am. J. Phys.* **61**, 329–335 (1993).
16. Rosendahl, J., Vekic, M. & Kelley, J. *Phys. Rev.* **E47**, 1401–1404 (1993).
17. Rosendahl, J., Vekic, M. & Rutledge, J. E. *Phys. Rev. Lett.* **73**, 537–540 (1994).
18. Feder, J. *Fractals* **3**, 431–433 (1995).

ACKNOWLEDGEMENTS. We thank N. Ainehchi, T. Engøy, S. Gundersen, U. Oxaal, R. J. Wallace and T. Walmann for their help at various stages of this project, and we also thank S. R. Nagel for discussions. This work was supported by VISTA, a research corporation between the Norwegian Academy of Science and Letters and Den norske stats oljeselskap a.s. (STATOIL), by NFR, the Research Council of Norway, and by Norsk Hydro a.s. K. C. was supported by The Danish Natural Science Research Council.

Giant magnetoresistance in $\text{Ti}_2\text{Mn}_2\text{O}_7$ with the pyrochlore structure

Y. Shimakawa, Y. Kubo & T. Manako

Fundamental Research Laboratories, NEC Corporation, 34 Miyukigaoka, Tsukuba 305, Japan

MATERIALS exhibiting giant magnetoresistance (GMR) undergo a large change in electrical resistance in response to an applied magnetic field. This effect is of technological interest as it can be exploited for the sensitive detection of magnetic fields in magnetic memory devices. A range of compounds have now been found to exhibit intrinsic GMR—these are all perovskites based on manganese oxide^{1–4}. Here we report the observation of GMR in $\text{Ti}_2\text{Mn}_2\text{O}_7$, which has a pyrochlore structure and thus differs both structurally and electronically from perovskites. At 135 K the magnetoresistance ratio (the change in resistance) reaches -86% at 7 tesla, comparable to the GMR response of perovskite materials. In contrast to the hole-doped perovskites, the charge carriers in our material are electrons, as determined from measurements of the Hall coefficient. The discovery of GMR in a second class of material expands the options for optimizing magnetoresistive properties for specific technological applications.

Giant magnetoresistance in the perovskite compounds originates from an applied magnetic field affecting spin-charge coupled magnetotransport phenomena. Hole-doping of the insulating parent compound LaMnO_3 (by substitution of Sr^{2+} for La^{3+} ions) produces Mn^{4+} ions, and thus creates itinerant e_g holes, which can hybridize with the $\text{O}(2p)$ state. A double-exchange interaction between Mn^{3+} and Mn^{4+} mediates ferromagnetic interaction, and the sample shows metallic conduction below the Curie temperature (T_C) (refs 5–7). Near T_C the applied magnetic field tends to align the local spins, and then electron transfer increases. As a result, a sharp resistivity decrease is observed.

Pyrochlore oxides with a general formula $\text{A}_2\text{B}_2\text{O}_7$ crystallize in the face-centred-cubic structure with eight formula units per unit cell and a space group of $Fd\bar{3}m$. Manganese ions in the pyrochlore $\text{Ti}_2\text{Mn}_2\text{O}_7$ occupy the 16c site in the structure and form slightly

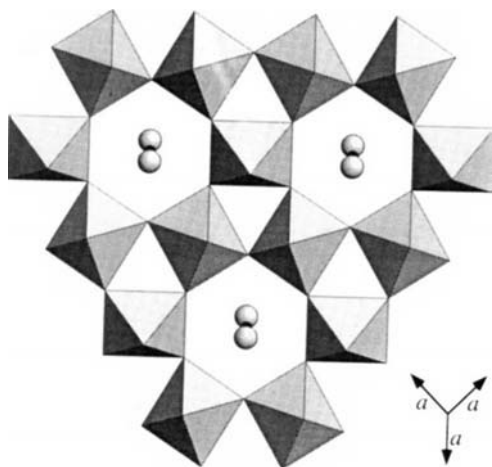


FIG. 1 A $\langle 111 \rangle$ projection of the pyrochlore structure showing the network of vertex-sharing MnO_6 octahedra. The Ti atoms (small black spheres) are in the centre of the hexagonal holes, with 8b oxygen atoms (larger, grey spheres) above and below. Cubic-lattice axes are shown with arrows marked a .

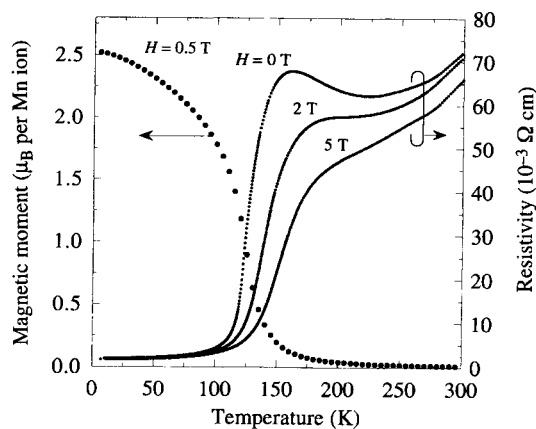


FIG. 2 Temperature dependence of magnetic moment measured with an applied magnetic field of 0.5 T. The temperature dependence of resistivity at 0, 2 and 5 T is also shown (right-hand axis).

distorted MnO_6 octahedra. $\text{Mn } 3d$ orbitals are split into t_{2g} and e_g states by the crystal field. Although both perovskite and pyrochlore manganese oxide structures consist of MnO_6 octahedra with vertex-sharing, these structures have different networks. In a pyrochlore structure, the MnO_6 octahedra form a three-dimensional array that has the composition MnO_3 or Mn_2O_6 with a Mn-O-Mn bond angle of $\sim 130^\circ$ (Fig. 1). A perovskite structure, in contrast, has a simple cubic network of MnO_6 with a nearly 180° bond angle.

Powder samples of $\text{Ti}_2\text{Mn}_2\text{O}_7$ (~ 500 mg) were synthesized by solid-state reaction. A mixture of Ti_2O_3 and MnO_2 was sealed in a gold capsule, and allowed to react in a high-pressure apparatus at 2.5 GPa at $1,020^\circ\text{C}$ for 30 min. The sample was rapidly cooled to room temperature before releasing the pressure. The cubic lattice constant, determined by a least-squares method from X-ray diffraction data, is $9.8966(4)$ Å.

Figure 2 shows the temperature dependence of the magnetic moment in an applied magnetic field of 0.5 T. The temperature dependence of the resistivity at 0, 2 and 5 T are also shown in the figure. The sample shows ferromagnetic behaviour at low temperatures. The susceptibility above 200 K obeys the Curie–Weiss law, $\chi = C/(T - \theta_p)$. T_C , determined by the temperature dependence of the ferromagnetic moment, is 142 K and the paramagnetic Curie temperature (θ_p) obtained from the Curie–Weiss plot is 164 K. Although the changes in magnetic moment and zero-field resistivity agree with those reported elsewhere^{8,9}, T_C and θ_p in the present experiment are slightly higher than the reported values. This may be caused by small differences in cation compositions and/or oxygen nonstoichiometry. The resistivity with zero applied field (ρ_{0T}) shows a broad maximum slightly above T_C and decreases markedly below this temperature (Fig. 2). The sharp decrease in resistivity in accordance with the development of spontaneous magnetization strongly suggests that the electron transfer is closely related to local spin alignment similar to the case with hole-doped perovskite manganese oxides.

Negative GMR effects are observed at around T_C . There is no difference in magnetoresistance (MR) effects for longitudinal (current I parallel to applied magnetic field H) and transverse (I perpendicular to H) configurations. At around T_C , the resistivity decreases dramatically when a magnetic field is applied. Figure 3 shows field-induced changes in resistivity at 120, 135 and 150 K. No saturation of the MR effect is observed in the range of the measured magnetic field. At 135 K, which is slightly lower than T_C , the MR ratio at 7 T reaches -86% . Such a huge ratio is comparable to the -90% change at 15 T observed in a single crystal of perovskite $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ (ref. 4).

Figure 4 shows the normalized resistivity change ρ/ρ_{0T} at 165 and 180 K plotted against normalized field-induced magnetization

M/M_s . Resistivity versus magnetization, which is obtained using $\rho(T)$ and $M(T)$ measured at 0.5 T, is also shown in the figure. Magnetoresistive changes at two temperatures (165 and 180 K) as well as the temperature dependence of resistivity lie on a single curve, suggesting a close relationship between changes in resistivity and magnetization. Magnetoresistance with field-induced magnetization (above T_C) and the sharp decrease in resistivity with the development of spontaneous magnetization (below T_C) are both governed by the same magnetic-ordering-related electron transfer mechanism.

In the small-magnetization region ($M \ll M_s$) of the perovskite manganese oxides above T_C , the MR magnitude is observed to obey the scaling function $-\Delta\rho/\rho_{0T} = C_{MR}(M/M_s)^2$, where C_{MR} is a coupling constant between the itinerant electrons and local spins^{4,10}. Weak-coupling theories have given the constants^{11,12}, but have failed to explain the experimental results. Furukawa recently investigated a ferromagnetic Kondo lattice model and obtained reasonable values of C_{MR} (ref. 13). However, the MR behaviour in the pyrochlore compound shown in Fig. 4 does not seem to obey this scaling function. When we fit the experimental data to the function in the very-small magnetization region, in which $(M/M_s)^2$ is less than 0.01, the obtained coupling constant C_{MR} is ~ 15 (Fig. 4 inset). This value is large compared with C_{MR} values of 1–4 in perovskite Mn oxides^{4,10}. Such discrepancies might result from the significant differences in electronic structures and carrier-doping mechanisms between the two compounds.

The observed magnetic moments, which originate from the Mn(3d) spin, indicate that the ionic state of manganese ions in $\text{Ti}_2\text{Mn}_2\text{O}_7$ is very close to that of Mn^{4+} ions. The saturated magnetic moment at 5 K is $2.59 \mu_B$ per Mn ion. The paramagnetic Curie constant C obtained from the Curie–Weiss plot gives an effective paramagnetic moment of $3.75 \mu_B$ per Mn ion. Expected moments for free Mn^{4+} ions are 3 ($S = 3/2$) and $3.87 \mu_B$ per Mn ion, respectively. Because the sample contains small amounts of impurities ($<10\%$), the observed moments should be slightly smaller than the intrinsic values. All these observations are consistent with a simple ionic consideration of the stoichiometric composition of $\text{Ti}_2\text{Mn}_2\text{O}_7$.

Measurements of Hall coefficients reveal that the predominant conduction carriers in $\text{Ti}_2\text{Mn}_2\text{O}_7$ are electrons. The observed Hall coefficients R_H at 260 and 315 K (above T_C) are approximately -6 to $-7 \times 10^{-1} \text{ cm}^3$ per coulomb. At temperatures below T_C , Hall resistivity changes linearly with magnetic fields up to 6 T and no anomalous Hall signal is observed. The R_H values observed at 10 and 50 K are about $-2 \times 10^{-1} \text{ cm}^3$ per coulomb. These Hall coefficients correspond to 0.001–0.005 conduction electrons per formula unit, and strongly indicate that a very small number of electrons are doped into the Mn^{4+} state. In the metallic perovskite

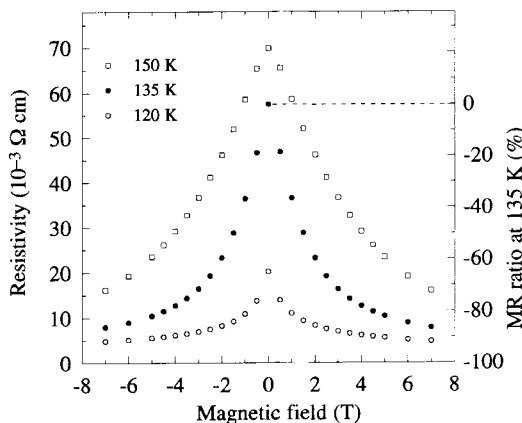


FIG. 3 Field-induced changes in resistivity at 120, 135 and 150 K. Magnetoresistance (MR) ratio $\Delta\rho/\rho_{0T}$ at 135 K is also shown (right-hand axis).

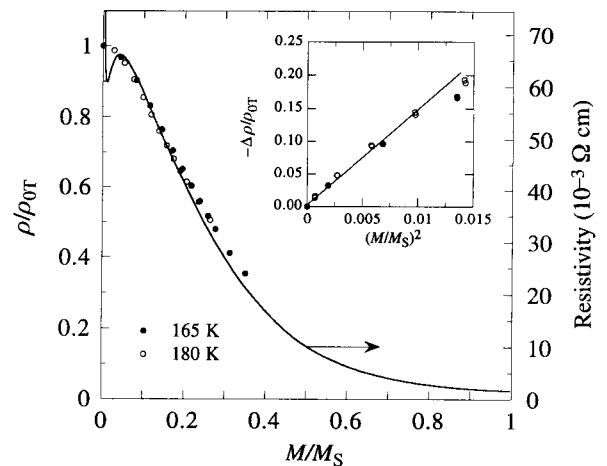


FIG. 4 Normalized resistivity ρ/ρ_{0T} at 165 (filled circles) and 180 K (empty circles) plotted against normalized field-induced magnetization M/M_s (M_s is the saturated magnetization). Resistivity versus magnetization, which is obtained using $\rho(T)$ and $M(T)$ measured at 0.5 T, is shown (curve; right-hand axis). The inset shows MR magnitudes $-\Delta\rho/\rho_{0T}$ plotted against the square of normalized field-induced magnetization $(M/M_s)^2$.

manganese oxides, in contrast, a fairly large number of holes (~ 0.18 holes per Mn ion) are introduced into the Mn^{3+} state in the parent insulator LaMnO_3 .

A possible explanation for the origin of the electron carriers is oxygen-deficiency, that is, a formula $\text{Ti}_2\text{Mn}_2\text{O}_{7-\delta}$. Oxygen atoms at an 8b site (Fig. 1) could be released from the structure, as in many kinds of oxygen-deficient pyrochlore compounds. A very small amount of oxygen vacancies can produce a sufficient number of conduction electrons in the system. The Ti(6s) band may also contribute to the transport properties because the energy level of the band is close to the levels of Mn(3d) and O(2p) bands.

We now consider two possible models of the electron doping into Mn^{4+} with a $(3d)^3$ configuration. As discussed by Goodenough¹⁴, exchange interaction of Mn ions (Δ_{ex}) and orbital splitting by the crystal field ($10Dq$) in octahedral-coordinated Mn^{4+} ions are very close. Hence, doped electrons are introduced into either the e_g or t_{2g} state depending on the energy difference between Δ_{ex} and $10Dq$. With a strong exchange interaction ($\Delta_{ex} > 10Dq$), the e_g state is favourable for the doped electrons (high-spin configuration $t_{2g}^3 e_g^1$). The e_g electrons behave itinerantly and (Hund) couple with the localized t_{2g} spins. This model is viewed as a symmetrical version of the hole-doped perovskite oxides. On the other hand, if $\Delta_{ex} < 10Dq$, extra electrons are introduced into the lower t_{2g} state (low-spin configuration t_{2g}^4). In that case the Fermi level crosses the bottom of the down-spin band of the t_{2g} state of Mn(3d), and thus the t_{2g} electrons travel on the Mn sites. In a band scheme, the Fermi surface consists of only the down-spin t_{2g} band, and the local moment originates from the fully occupied up-spin t_{2g} band. A preliminary local density approximation (LDA) calculation predicts the latter possibility (N. Hamada, personal communication).

In the perovskite double-exchange systems, interaction between the Mn ions in the parent insulator is antiferromagnetic. Kinetic energy gain by doped carriers stabilizes the ferromagnetic alignment of the Hund-coupled Mn moments^{5,7}. Because the conduction carriers are so small in the pyrochlore, $\text{Ti}_2\text{Mn}_2\text{O}_7$, an ambiguity remains as to whether the double-exchange interaction is the only mechanism for stabilizing the ferromagnetic and metallic state. Considering that some manganese pyrochlore insulators such as $\text{Y}_2\text{Mn}_2\text{O}_7$ and $\text{Lu}_2\text{Mn}_2\text{O}_7$ show ferromagnetism at low temperatures¹⁵, the ferromagnetic interaction via a super-exchange-type interaction may also occur at low temperatures. Thus, the double-exchange interaction may play a role in enhancing the ferromagnetic interaction. □

Received 30 August; accepted 13 November 1995.

- Chahara, K., Ohno, T., Kasai, M. & Kozono, Y. *Appl. Phys. Lett.* **63**, 1990–1992 (1993).
- von Helmolt, R., Wecker, J., Holzapfel, B., Schultz, L. & Samwer, K. *Phys. Rev. Lett.* **71**, 2331–2333 (1993).
- Jin, S. et al. *Science* **264**, 413–415 (1994).
- Tokura, Y. et al. *J. phys. Soc. Japan* **63**, 3931–3935 (1994).
- Jonker, G. H. & van Santen, J. H. *Physica* **16**, 337–349 (1950).
- Zener, C. *Phys. Rev.* **82**, 403–405 (1951).
- de Gennes, P.-G. *Phys. Rev.* **118**, 141–154 (1960).
- Fujinaka, H., Kinomura, N., Koizumi, M., Miyamoto, Y. & Kume, S. *Mater. Res. Bull.* **14**, 1133–1137 (1979).
- Raju, N. P., Greedan, J. E. & Subramanian, M. A. *Phys. Rev. B* **49**, 1086–1091 (1994).
- Urushibara, A. et al. *Phys. Rev. B* **51**, 14103–14109 (1995).
- Kubo, K. & Ohata, N. *J. phys. Soc. Japan* **33**, 21–32 (1972).
- Searle, C. W. & Wang, S. T. *Can. J. Phys.* **48**, 2023–2031 (1970).
- Furukawa, N. *J. phys. Soc. Japan* **63**, 3214–3217 (1994).
- Goodenough, J. B. in *Progress in Solid State Chemistry* Vol. 5 (ed. Reiss, H.) 145–399 (Pergamon, Oxford, 1971).
- Subramanian, M. A., Torardi, C. C., Johnson, D. C., Pannetier, J. & Sleight, A. W. *J. Solid St. Chem.* **72**, 24–30 (1988).

ACKNOWLEDGEMENTS. We thank N. Hamada for helpful discussions and an LDA calculation for $\text{Ti}_2\text{Mn}_2\text{O}_7$.

Hydrogen-bond kinetics in liquid water

Alenka Luzar & David Chandler

Department of Chemistry, University of California, Berkeley, California 94720-1460, USA

HYDROGEN bonds play a crucial role in the behaviour of water^{1–4}; their spatial patterns and fluctuations characterize the structure and dynamics of the liquid^{5–7}. The processes of breaking and making hydrogen bonds in the condensed phase can be probed indirectly by a variety of experimental techniques⁸, and more quantitative information can be obtained from computer simulations⁹. In particular, simulations have revealed that on long timescales the relaxation behaviour of hydrogen bonds in liquid water exhibit non-exponential kinetics^{7,10–13}, suggesting that bond making and breaking are not simple processes characterized by well defined rate constants. Here we show that these kinetics can be understood in terms of an interplay between diffusion and hydrogen-bond dynamics. In our model, which can be extended to other hydrogen-bonded liquids, diffusion governs whether a specific pair of water molecules are near neighbours, and hydrogen bonds between such pairs form and persist at random with average lifetimes determined by rate constants for bond making and breaking.

In a computer model of liquid water, an instantaneous configuration, $r(t)$, denotes the positions of all the atoms in the system at time t . A configurational criterion for whether a particular pair of water molecules is bonded allows one to define a hydrogen-bond population operator, $h[r(t)] = h(t)$. This operator has a value 1 when the particular tagged pair, say molecules 1 and 2, are bonded, and zero otherwise. The average number of hydrogen bonds in an equilibrium fluid of N water molecules is $\frac{1}{2}N(N-1)\langle h \rangle$, where $\langle h \rangle$ denotes the average of $h[r(t)]$.

In the dynamic equilibrium of liquid water, hydrogen-bond populations fluctuate in time. The fluctuations are characterized by the correlation function

$$c(t) = \langle h(0)h(t) \rangle / \langle h \rangle \quad (1)$$

This function is the probability that the hydrogen bond is intact at time t , given that it was intact at time zero. It is computed from a simulation trajectory by recording the joint occurrences of two non-zero populations separated by a time t . At equilibrium, the probability that a specific pair is bonded in a large system is negligibly small. As a result, $c(t)$ relaxes to zero. According to the principles of statistical mechanics, specifically the fluctuation-

dissipation theorem¹⁴, this time evolution of $c(t)$ is the same as that for an initially prepared non-equilibrium concentration of hydrogen bonds. The rate of relaxation to equilibrium is

$$k(t) = -dc/dt = -\langle \dot{h}(0)[1 - h(t)] \rangle / \langle h \rangle \quad (2)$$

where $\dot{h}(0) = (dh/dt)_{t=0}$. Note that, because $\langle h(0)\dot{h}(t) \rangle = -\langle \dot{h}(0)h(t) \rangle$ and $\langle \dot{h}(0) \rangle = 0$, equation (2) follows from equation (1). Thus $-k(t)$ is the average rate of change of hydrogen-bond population for those trajectories where the bond is broken at a time t later. Figure 1 shows the $k(t)$ determined from our simulations of water at room temperature.

At short times, $k(t)$ manifests an assortment of motions leading to hydrogen-bond breaking. The most important of these motions are librations on the timescale of less than 0.1 ps, and inter-oxygen vibrations on the timescale of 0.1–0.2 ps. Beyond this transient period, $k(t)$ decays monotonically. To the extent that each hydrogen bond acts independently of other hydrogen bonds and also independently of other processes of similar timescales, the long time decay of $k(t)$ would be that of first-order kinetics^{15,16}. That is, one would expect $k(t) \approx v \exp(-vt)$, where $1/v$ would be the average hydrogen-bond lifetime. But as seen in Fig. 1, the post-transient relaxation of $k(t)$ is decidedly different. Beyond the transient, the slope of $\log k(t)$ increases monotonically with time t . The log-log plot shows clearly that this behaviour does not coincide with a power-law decay.

The distribution of hydrogen-bond lifetimes provides a complementary view of hydrogen-bond dynamics. This distribution is determined from trajectory calculations by recording the separate lengths of time over which $h[r(t)] = 1$. The functions $c(t)$ and $k(t)$ would decay exponentially if, and only if, the lifetime distribution were also exponential. The lifetime distribution is, however, non-exponential¹¹. Over a limited range of times, it exhibits power-law

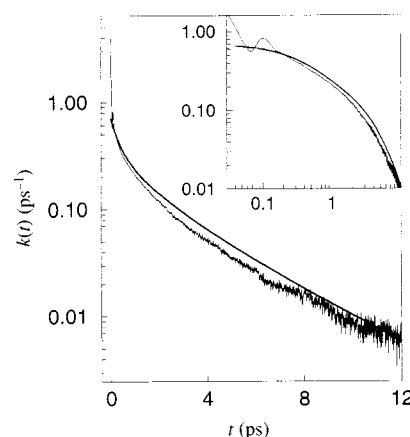


FIG. 1 The rate function, $k(t)$, computed from the trajectories of a molecular-dynamics computer simulation (thin line) and from the diffusion model (thick line). The simulation was for water at room temperature, and with a density of 1 g cm^{-3} . The length of the trajectory coincided with 240 ps of physical time. It employed the SPC potential model for water²³, using periodic boundary conditions and Ewald sums²⁴. In analysing the trajectory, two water molecules are considered to be hydrogen-bonded only if their inter-oxygen distance is $< 3.5 \text{ \AA}$, and simultaneously the angle between the O–O axis and one of the O–H bonds is 30° (ref. 19). There exists an assortment of reasonable alternative choices corresponding to different surfaces in configuration space dividing bonded from unbonded regions. Because trajectories flow quickly between nearby dividing surfaces, however, the behaviour of $k(t)$ is sensitive to hydrogen-bond definition for short times only. Beyond the transient period, the behavior of $k(t)$ is invariant to physically reasonable changes in hydrogen-bond definition. The main figure is a semi-log plot, the inset is a log-log plot. For the diffusion model, $k(t)$ is the inverse Laplace transform of $\bar{k}(s) = k/[s + k + k'sf(s)]$, where $f(s) = 3\tau[1 - \sqrt{s\tau} \arctan(1/\sqrt{s\tau})]$. These are the relationships obtained from the solution of equations (5) and (6) in the text. The inverse Laplace transform is evaluated by numerical integration. The parameters used in the plots are $\tau = 0.4 \text{ ps}$, $k = 0.7 \text{ ps}^{-1}$ and $k' = 1.0 \text{ ps}^{-1}$.

behaviour¹¹. These dynamic properties of a single tagged hydrogen bond should not be confused with that of the total population of hydrogen bonds. Because the total number of hydrogen bonds is extensive, their fluctuations are relatively small¹⁴, and the correlation function for the total number of hydrogen bonds appears exponential on the timescale of picoseconds (ref. 17).

It is unlikely that the source of non-exponential kinetics of tagged hydrogen bonds in water is due to correlations between different hydrogen bonds. The lifetime distribution that exhibits a power law for only a limited range of times suggests that correlations do not persist over large lengths and thus do not involve many hydrogen bonds. Further, the apparent exponential kinetics of the total hydrogen-bond population suggests that this collective variable is a sum of many statistically independent entities. Most importantly, however, by partitioning the trajectories that contribute to $k(t)$ according to the density of hydrogen bonds that are neighbours to a tagged bond, we have found that the dependence at room temperature is insignificant (A.L. and D.C., unpublished data).

A more likely source of significant non-exponential relaxation at ambient conditions is the coupling between hydrogen-bond population and diffusion. Two bonded molecules can diffuse apart only if the hydrogen bond between them breaks. Further, a broken hydrogen bond can reform if a molecule reverses its direction and diffuses back to its partner. To examine the quantitative consequences of diffusion, we have partitioned the contributions to $k(t)$ according to whether or not the pair has moved apart after its bond has broken. From this partitioning, we have computed

$$k_{in}(t) = -\langle \dot{h}(0)[1 - h(t)]H(t) \rangle / \langle h \rangle \quad (3)$$

Here $H(t)$ is unity if the oxygen–oxygen distance of the tagged pair is not larger than some distance $R = 3.5 \text{ \AA}$, and it is zero otherwise. In the hydrogen-bond definition we employ^{18,19}, two water molecules separated by less than $3.5 \text{ \AA}$ can be either bonded or not bonded, depending upon their relative orientations. At larger separations, a bond cannot form. The first coordination shell of water, as measured by its oxygen–oxygen radial distribution function²⁰, extends to $\sim 3.5 \text{ \AA}$.

The probability at time t that a pair of initially bonded water molecules are now unbonded but remain separated by less than R is

$$n(t) = \int_0^t dt' k_{in}(t') \quad (4)$$

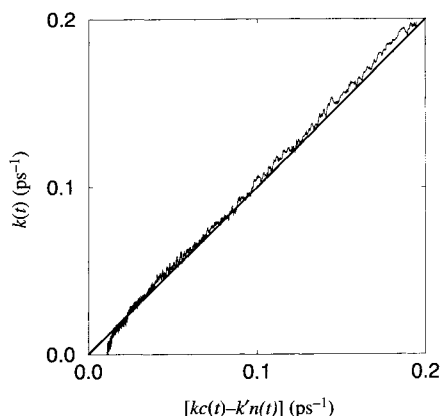


FIG. 2 Correlation plot relating the functions $c(t)$, $n(t)$ and $k(t)$ (see text), obtained from the molecular-dynamics computer simulation, for times t from 1 to 20 ps. At shorter times, deviations from the correlation are large. At longer times, statistical uncertainties of the simulation results preclude a test of the correlation. The validity of equation (5) is judged by comparing the simulation data (jagged line) with the straight line of unit slope. A description of the simulation is given in Fig. 1 legend. The values of the constants k and k' used in the plot are 0.7 ps^{-1} and 1.0 ps^{-1} , respectively.

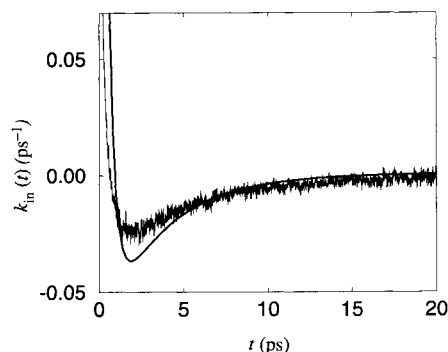


FIG. 3 The restricted rate function, $k_{in}(t)$, as computed from trajectories of a molecular-dynamics simulation (thin line) and from the diffusion model (thick line). A description of the computer simulation is given in Fig. 1 legend. For the diffusion model, $k_{in}(t)$ is the inverse Laplace transform of $k_{in}(s) = sf(s)k(s)$, where $k(s)$ and $f(s)$ are given in the Fig. 1 legend. The inverse Laplace transform is evaluated by numerical integration. The parameters used in the plot are $\tau = 0.4 \text{ ps}$, $k = 0.7 \text{ ps}^{-1}$ and $k' = 1.0 \text{ ps}^{-1}$.

The probabilities $c(t)$ and $n(t)$ correspond to local populations that can interconvert. Hence, their kinetics might follow

$$dc/dt = -kc(t) + k'n(t) \quad (5)$$

where k and k' (not to be confused with the function $k(t)$) are the respective rate constants for the breaking and making of hydrogen bonds between a near-neighbour pair of water molecules. The correlation plot given in Fig. 2 shows that with $k = 0.7 \text{ ps}^{-1}$ and $k' = 1.0 \text{ ps}^{-1}$ equation (5) is consistent with the results of our computer simulation for times beyond the transient period. If k and k' differ by more than 10% from these values, the simulation data deviate significantly from the line of unit slope. The relaxation rate constants thus identified for hydrogen bonds in water are material properties. For example, the values of k and k' depend on the temperature and concentration of solutes in the system, and will also be sensitive to H/D isotopic variation. Similarly, for simulations, the values of k and k' can depend on the model used. Irrespective of the system, however, equation (5) provides the connection between rigorous microscopic dynamics and hydrogen-bond rate constants.

The physical meaning of $1/k$ is that of the average hydrogen-bond lifetime. The value obtained from the correlation plot is 1.4 ps. Essentially the same average lifetime is obtained from the distribution of lifetimes in our simulation, as should be the case if equation (5) is correct. In the broken bond state, the bond between the tagged pair of molecules, 1 and 2, no longer exists as these molecules are now bonded to other nearby partners. The average frequency of switching allegiances is $k + k'$, which is 1.7 ps^{-1} in the simulation model.

To the extent that $n(t)$ is non-zero, the two tagged molecules are no longer bonded to each other but remain within a distance R . The population $n(t)$ is therefore a measure of local strain in the hydrogen bonds. It will relax not only by conversion back to the bonded state, but also by diffusion. If this diffusion were to occur slowly enough, $c(t) + n(t)$ would appear to be a constant. The separate populations, $c(t)$ and $n(t)$, would each change by interconversion only, relaxing as single exponentials. But diffusion can occur on a time scale that is comparable to $1/k$ and $1/k'$.

The degree to which diffusion is significant can be estimated from

$$\frac{\partial}{\partial t} \rho(\mathbf{r}, t) = D \nabla^2 \rho(\mathbf{r}, t) + \delta(\mathbf{r}) [kc(t) - k'n(t)] \quad (6)$$

Here $\rho(\mathbf{r}, t)$ represents the density of the diffusing unbonded pair, where $\mathbf{r}$ is the vector between the pair, and D is the interdiffusion constant for the pair. Equation (6) is restricted to timescales larger than that of the transient interval, and to length scales larger than the spatial width of a hydrogen bond. Greater resolution would

require changing k , k' and D from constants to operators that are non-local in space and time. The delta function of the source and sink terms in equation (6) localizes $\mathbf{r}$ to within a volume a^3 , where a is the range of lengths over which a hydrogen bond might exist. The quantity $\rho(\mathbf{0}, t)$ is thus $n(t)/a^3$. This connection, with equations (5) and (6) gives three equations for three unknowns, and these comprise a diffusion model for hydrogen-bond kinetics. The model can be solved straightforwardly for $k(t)$ and $k_{in}(t)$. The solution depends on the length a as it appears in the time $\tau = a^2/D(6\pi^2)^{2/3}$. Diffusion constants of water are of the order of $10^{-5} \text{ cm}^2 \text{ s}^{-1}$. The radius of a water molecule is $\sim 1.5 \text{ \AA}$. Thus, τ is expected to be in the range of 0.1–1 ps, neither short nor long compared to $1/k$ and $1/k'$.

Figure 1 shows a comparison of $k(t)$ obtained from our computer simulation with that obtained from the diffusion model with $\tau = 0.4 \text{ ps}$ and with k and k' fixed at the values found from the correlation plot, Fig. 2. Less favourable comparisons are found by varying τ by 25% from this value. A similar comparison using the same values of k , k' and τ is made in Fig. 3 for $k_{in}(t)$. The diffusion model is seen to provide an explanation of most of what is found from the computer simulation calculations, including the fact that $k_{in}(t)$ becomes negative after a period of somewhat less than 1 ps. This behaviour is associated with trajectories that initially have a $\dot{h}(0) > 0$, and thus initially enter into a hydrogen bond. These trajectories are trapped, re-emerging only later into the unbonded region of configuration space.

Quantitative improvements to the diffusion model could be made by refining its resolution in space and time. Independent of such improvements, our analysis establishes the existence of well defined rate constants k and k' and their connection to microscopic correlation functions. This connection can be used in general to analyse the dynamics of hydrogen-bonded liquids. Our phenomenological model can be applied to interpret inco-

herent neutron scattering²¹ and perhaps transient vibrational spectroscopy²². Experimental verification of the model will require coordinated application of two different techniques because $c(t)$ and $n(t)$ are independent functions that vary on the same timescale. $\square$

Received 26 September; accepted 13 November 1995.

1. Teixeira, J. J. *Physique IV* C1, **3**, 162–169 (1993).
2. Eisenberg, D. & Kauzmann, W. *The Structure and Properties of Water* (Oxford Univ. Press, New York, 1969).
3. Franks, F. (ed.) *Water Science Reviews* Vols 1–5 (Cambridge Univ. Press, 1985–90).
4. Stanley, H. E. & Ostrowsky, N. (eds) *Correlations and Connectivity, Geometric Aspects of Physics, Chemistry and Biology* (Kluwer Academic, Dordrecht, 1990).
5. Stillinger, F. H. *Adv. chem. Phys.* **31**, 1–101 (1975).
6. Stillinger, F. H. *Science* **209**, 451–457 (1980).
7. Ohrnime, I. & Tanaka, H. *Chem. Rev.* **93**, 2545–2566 (1993).
8. Dore, J. C. & Teixeira, J. (eds) *Hydrogen-Bonded Liquids* (Kluwer Academic, Dordrecht, 1991).
9. Ladanyi, B. M. & Skaf, M. S. A. *Rev. Chem.* **44**, 335–368 (1993).
10. Belch, A. C. & Rice, S. A. *J. chem. Phys.* **86**, 5676–5682 (1987).
11. Sciortino, F., Poole, P. H., Stanley, H. E. & Havlin, S. *Phys. Rev. Lett.* **64**, 1686–1689 (1990).
12. Zichi, D. A. & Rossky, P. J. *J. chem. Phys.* **84**, 2814–2822 (1986).
13. Luzar, A. & Chandler, D. in *Hydrogen Bond Networks* (eds Bellissent-Funel, M. C. & Dore, J. C.) 239–246 (Kluwer Academic, Dordrecht, 1994).
14. Chandler, D. *Introduction to Modern Statistical Mechanics* (Oxford Univ. Press, New York, 1987).
15. Chandler, D. *J. chem. Phys.* **68**, 2959–2970 (1978).
16. Berne, B. J. in *Multiple Time Scales* (eds Brackbill, J. U. & Cohen, B. I.) 419–436 (Academic, New York, 1985).
17. Saito, S. & Ohrnime, I. *J. chem. Phys.* **102**, 3566–3579 (1995).
18. Ferrario, M., Haughey, M., McDonald, I. R. & Klein, M. L. *J. chem. Phys.* **93**, 5156–5166 (1990).
19. Luzar, A. & Chandler, D. *J. chem. Phys.* **98**, 8160–8173 (1993).
20. Soper, A. K. & Phillips, M. G. *Chem. Phys.* **107**, 47–60 (1986).
21. Teixeira, J., Bellissent-Funel, M.-C., Chen, S. H. & Dianoux, A. J. *Phys. Rev. A* **31**, 1913–1917 (1985).
22. Bratos, S. & Leicknam, J.-C. *J. chem. Phys.* **103**, 4887–4893 (1995).
23. Berendsen, H. J. C., Postma, J. P. M., van Gasteren, W. F. & Hermans, J. in *Intermolecular Forces* (ed. Pullman, B.) 331–342 (Reidel, Dordrecht, 1981).
24. Allan, M. P. & Tildesley, D. J. *Computer Simulation of Liquids* (Clarendon, Oxford, 1987).

ACKNOWLEDGEMENTS. We thank E. A. Chandler for suggesting the use of equation (6), and H. E. Stanley for comments on the earlier draft of the manuscript. The work was supported by the US ONR.

Dynamics of CO₂-driven lake eruptions

Youxue Zhang

Department of Geological Sciences, The University of Michigan, Ann Arbor, Michigan 48109-1063, USA

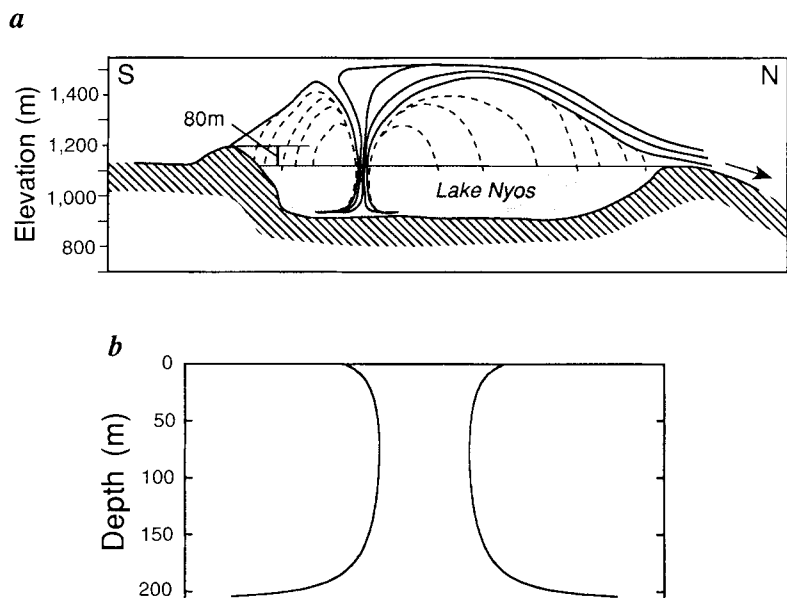
ON 21 August 1986, a massive release of carbon dioxide from Lake Nyos in Cameroon killed about 1,700 people. A similar event occurred on 15 August 1984 at Lake Monoun, also in Cameroon. It was suggested^{1–5} that the CO₂ released was initially dissolved in the hypolimnion (dense lower layer) of the lake, and was released by eruptive outgassing. Because of its violence, the Nyos outburst was at first thought⁶ to have been volcanic, but undisturbed sediments and other evidence indicate that no large volcanic eruption occurred^{7–9}. Recent experiments^{10,11} have shown that decompression of CO₂-saturated water is able to power explosive eruptions. Here I analyse the dynamics of CO₂-driven lake-water eruptions by deriving an equation of state for gas–liquid mixtures and using it to integrate the Bernoulli equation, which describes the dynamics of the bubbly flow. I find that under certain conditions these eruptions can be violent: the lake-surface exit velocity of an initially gas-saturated water parcel may reach 89 m s^{-1} for Lake Nyos and 51 m s^{-1} for Lake Monoun. The dynamics are similar to those of water-driven volcanic eruptions, which are also powered by gas exsolution from a liquid.

The dissolved CO₂ content of the water of Lake Nyos, and hence the density of that water, increases with depth owing to slow leakage of magmatic CO₂ into the lake^{8,9}. But as the partial pressure of CO₂ (p_{CO_2}) in the water increases, the hydrostatically

stable water at the lake bottom becomes increasingly dynamically unstable because a small perturbation due to a landslide, sinking of cold rain water, an internal wave, a small volcanic injection of CO₂, or a heavy flood of water into the lake might move the water up sufficiently to reach saturation ($p_{\text{CO}_2} = P_{\text{total}}$), allowing bubbles to form and grow. Being less dense than the surrounding water, the bubbly water ascends with increasing speed, resulting in an “overturn” of the lake or a limnic eruption¹, which releases the lethal CO₂ (ref. 9). The dynamics of CO₂-driven limnic eruptions have not been fully investigated¹².

Rise of CO₂-saturated water may be explosive because of strong positive feedback between bubble formation and growth, volume expansion and buoyancy rise. Once triggered, uprising of bottom water may form an erupting “conduit”^{13,14} into which more bottom water may be drawn (Fig. 1), allowing it to sustain itself throughout the eruption. In the Lake Nyos eruption, an estimated $(1\text{--}3) \times 10^8 \text{ kg}$ of CO₂ (refs 5, 9) was released during an eruption that possibly lasted as long as 4.5 hours (ref. 12). Hence the process probably lasted much longer than the time needed for a water parcel to ascend through the erupting conduit ($\leq 10 \text{ s}$, see later discussion on ascent velocity). That is, the process may be roughly at steady-state. Steady state also requires that the pressure in the conduit at each depth be independent of time. Because the medium through which limnic eruptions occur is fluid, the pressure P at each depth in the erupting conduit is roughly hydrostatic (denoted as P_h) and hence independent of time. Otherwise, water surrounding the conduit would flow into, compress and constrict the conduit so as to maintain hydrostatic equilibrium if the pressure in the conduit at each depth is less than P_h ; surrounding water would flow away to let the conduit expand if the pressure inside the conduit is greater than P_h . The approximate independence of pressure with time shows that the

FIG. 1 a, Schematic south–north cross-section of Lake Nyos showing probable movement of gas cloud (solid curves) and water droplets (dashed curves) during a limnic eruption driven by CO_2 dissolved in the water. The trajectory of water droplets depends on their size; larger droplets rain down closer to the conduit. The vertical and the horizontal scales are the same. Turbidity of the flow is not shown. The position of the conduit is not important to the model presented and could as well be towards the north end of the lake¹. The flow is shown as non-symmetrical because of possible wind. b, Close-up view of idealized erupting conduit shape, calculated by assuming that the mass flux is constant; that is, $A\rho u = \text{constant} = (A\rho u)_{\text{exit}}$ where A is the cross-section area. The conduit diameter depends on the mass flux into it (hence no scale is given for the horizontal axis) and reaches a minimum at a depth of 77 m.



flow can be steady once initiated and when there is enough supply of CO_2 -saturated (or nearly saturated) water. Therefore, the flow is assumed to be roughly at steady-state in the model presented below.

Denoting the depth below water surface as h , the total lake depth as h_0 , and the height above the bottom as z , then $z = h_0 - h$. In an erupting conduit, the pressure increases with depth as $dP = \rho(g + a)dh = -\rho(g + a)dz$, where ρ is the density of the gas–liquid mixture and varies with pressure (compressible flow), g is acceleration due to gravity, and a is the upward acceleration of the flow and varies with z . Thus $d(u^2/2)/dz = a = -g - (1/\rho)dP/dz$ where u is the ascent velocity depending on z . Furthermore, $P \approx P_h = P_0 - \rho_0 gz = P_{\text{atm}} + \rho_0 gh$, where ρ_0 is the density of bubble-free water and is assumed to be constant, and P_{atm} is the atmospheric pressure on the surface of the lake. If the flow is also barotropic (that is, local ρ is a function of P only), integration of $d(u^2/2) = -gzdz - (1/\rho)dP$ from $z = 0$ (where $P = P_0$ and $u = 0$) to an arbitrary z (where pressure is P) leads to the Bernoulli equation:

$$\int_{P_0}^P \frac{dP}{\rho} + \frac{1}{2}u^2 + gz = 0 \quad (1)$$

If the saturation depth is above lake bottom, the pressure at that depth should be substituted for P_0 in equation (1) and the equations given below. To integrate equation (1), the relationship between ρ and P , that is, the equation of state for the gas–liquid mixture, must also be known. The liquid density is roughly constant (ρ_0), and the liquid mass is also roughly constant. Hence, $\rho_0/\rho \approx 1 + V_{\text{gas}}/V_{\text{liq}}$ where $V_{\text{gas}}/V_{\text{liq}}$ is the ratio of the local gas-phase volume to liquid volume. The decompression process is roughly adiabatic with respect to the gas–liquid mixture. Because of the high heat capacity of water, the temperature decrease due to gas exsolution and expansion is small (the temperature decrease is estimated as $\sim 4^\circ\text{C}$ if P_0 is 2.0 MPa and final P is 0.1 MPa). Therefore the effect of temperature variation on density is ignored. When the pressure is less than 2 MPa, the CO_2 gas is roughly ideal. Experiments show that in a CO_2 –water system bubble growth is rapid (millimetre-sized bubbles can form in 0.02 s) and nucleation of bubbles is not difficult even with distilled water^{10,11}. Hence, details of nucleation and bubble growth are not considered and bubble growth is assumed to be rapid enough to keep equilibrium between the gas and liquid. This assumption is different from that of Wilson¹⁵ who assumed that for volcanic eruptions the mass of the gas phase is constant after fragmentation. The difference is due to the much greater diffusivity of CO_2 in water at room temperature¹⁶ than that of H_2O in

rhyolitic magma at 850°C (ref. 17). For example, the diffusion distance of CO_2 in water is $45\ \mu\text{m}$ in one second whereas that of H_2O in magma is $\sim 3\ \mu\text{m}$ in one second. Fragmented water droplets are expected to be small and diffusion may be rapid enough to maintain rough chemical equilibrium.

For CO_2 in water below 2 MPa, the Ostwald solubility coefficient λ is roughly independent of pressure. With the above assumptions and approximations, $\lambda \equiv C_{\text{liq}}/C_{\text{gas}} = (n_{\text{liq}}/V_{\text{liq}})/[(n_0 - n_{\text{liq}})/V_{\text{gas}}] \approx (V_{\text{gas}}/V_{\text{liq}})/(P_0/P - 1)$ where n_{liq} is the number of moles of the gas in the liquid, C_{liq} is the gas concentration in the liquid in mol m^{-3} and C_{gas} is that in the gas. Therefore $V_{\text{gas}}/V_{\text{liq}}$ can be replaced by $\lambda(P_0/P - 1)$ and the equation of state for a CO_2 –water mixture becomes:

$$\rho_0/\rho \approx 1 + \lambda P_0/P - \lambda \quad (2)$$

Using the above equation to integrate $\int (1/\rho)dP$ in the Bernoulli equation yields:

$$\frac{1}{2}u^2 \approx \lambda \frac{P_0}{\rho_{\text{liq}}} \left(\ln \frac{P_0}{P} - 1 + \frac{P}{P_0} \right) \quad (3)$$

The acceleration of the upward flow can be found to be $\lambda g(P_0/P - 1)$ in which P is decreasing. That is, the ascending flow has an increasing acceleration owing to continuous pressure decrease, unlike the constant acceleration in experiments in which CO_2 -saturated water is decompressed instantaneously to a constant pressure^{10,11}. The velocity u , calculated using equation (3), is a maximum estimate because the effects of turbulence, shallow water entrainment, and oversaturation of CO_2 in water have been ignored. Hence, a more accurate account of the non-ideality of CO_2 , the variation of λ with pressure, and the density variation of CO_2 -saturated water is not warranted at present.

Figure 2a shows u as a function of depth for Lake Nyos assuming an initially saturated water parcel (with 3 wt% of dissolved CO_2) starts at lake bottom. As expected, the velocity increases as the gas–liquid mixture rises. With decreasing depth, the velocity first increases slowly and then rapidly. As the column leaves the lake surface at 0.1 MPa pressure, the maximum exit velocity of the erupting column is $89\ \text{ms}^{-1}$. At this maximum exit velocity and ignoring entrainment of air, the erupting column would rise to a height of 400 m ($0.5u_{\text{exit}}^2/g$). The minimum column height was estimated to be 120 m for the 1986 Lake Nyos eruption based on the distribution of dead animals⁵.

Figure 2b shows the calculated maximum exit velocity as a function of saturation depth. For Lake Monoun (depth 96 m), the calculated maximum exit velocity is $\sim 51\ \text{m s}^{-1}$. Ignoring entrain-

ment of air, an erupting column with such velocity would rise to a height of ~ 130 m. If Lake Nyos erupted from a saturation depth of 150 m (instead of 208 m), the maximum exit velocity would be 70 m s^{-1} and the column would rise to a height of ~ 250 m. The above results, with modifications to account for different initial and boundary conditions, may also be applied to analyse the dynamics of eruption through pipes designed to degas the gas-saturated deep water in these lakes¹⁸.

Magmatic CO_2 leaks into the bottom of lakes Nyos and Monoun at a rate that could cause saturation of water at lake bottom on a timescale of years to decades¹⁸. It has been suggested that the 1986 eruption of Lake Nyos began at the base of a shallow chemocline^{13,14}. Although eruption triggering is stochastic and could occur at shallower depths, a more disastrous, more easily triggered, and possibly more likely event would be the eruption of a significant amount of CO_2 -rich bottom water if a substantial thickness of the bottom water is near saturation. Therefore, I present below and in Fig. 1 a model for a limnic eruption sustained by saturated bottom water. Because of some triggering mechanism, saturated bottom water is displaced upward. An erupting conduit is formed and sustained by drawing more bottom water into the rapidly rising column. By analogy with experimental results^{10,11}, the rising flow is initially roughly a uniform bubbly flow and then fragments into a mist⁹ when porosity reaches some

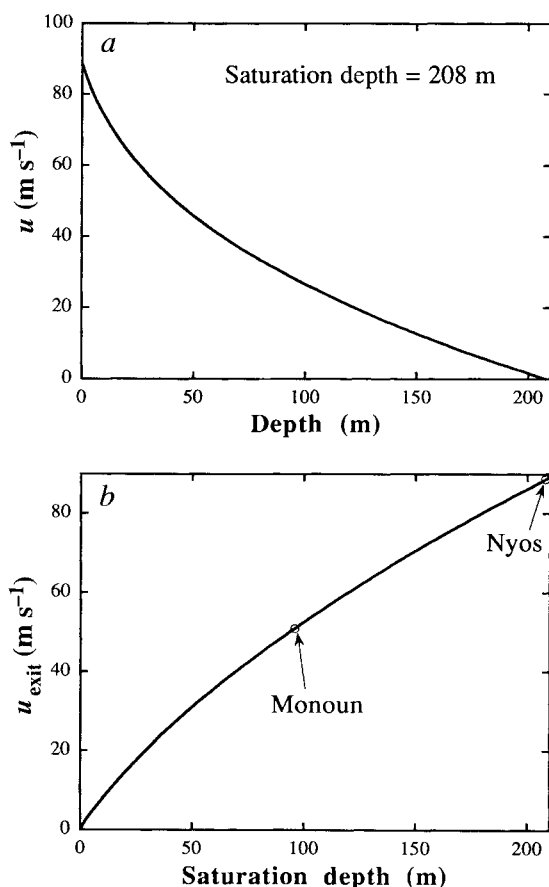


FIG. 2 a, Calculated maximum ascent velocity (u) of an erupting CO_2 -water mixture as a function of depth for Lake Nyos. The calculation uses equation (3) with $\lambda = 0.87$ ($T = 23^\circ\text{C}$), $P_0 = 2.14 \text{ MPa}$, $\rho_0 = 1,000 \text{ kg m}^{-3}$, and pressure P in the conduit being hydrostatic. The λ value is calculated using the equation $\ln \lambda = 3.12272 - 4042/T + 911269/T^2$ (where T is in kelvin), derived from data in refs 20–23. The value of u is not strongly dependent on small variations of the atmospheric pressure. For example, changing the atmospheric pressure from 0.1 to 0.088 MPa (for an elevation of 1,100 m) changes u_{exit} from 89 to 91 m s^{-1} . b, Calculated maximum exit velocity (u_{exit}) for CO_2 -driven water eruption as a function of initial saturation depth using equation (3) with $\lambda = 0.87$ and $P = 0.1 \text{ MPa}$. The u_{exit} for the eruption of bottom water of lakes Nyos and Monoun is indicated by arrows.

critical value (70–80%). Water droplets are carried up by the ascending gas and are roughly in thermal and chemical equilibrium with the gas, depending on the size of the droplets. Therefore, not only is CO_2 gas lost from the lake, but water would rise above the surface level of the lake. Most of the water probably rains back into the lake. Because CO_2 -rich gas is denser than air, the erupting column eventually collapses to form a CO_2 flow down the external slope (a gravity current), carrying fine water droplets with it. By analogy to ‘pyroclastic’ flow, the CO_2 flow carrying water droplets at ambient temperature may be termed an ambioructic (combination of ‘ambient’ and ‘eruct’) flow. At the exit velocity of 89 m s^{-1} , a conduit ~ 10 m in diameter could erupt CO_2 at a rate ($4 \times 10^7 \text{ kg h}^{-1}$) comparable to that estimated for the 1986 Lake Nyos eruption^{5,9,12}.

Tazieff⁶, assuming that a lake overturn or a limnic eruption would be slow, used the violence and localized nature of the process to argue against the limnic hypothesis and to suggest the volcanic hypothesis. My calculations and arguments above show that CO_2 -driven water eruptions can be violent (as also shown by experiments^{10,11}) and localized. Therefore, the Lake Nyos and Lake Monoun events are consistent with CO_2 -driven water eruption. Because of the large amount of CO_2 dissolved in the deep water of Lake Nyos, any volcanic-gas eruption through the lake would probably trigger a CO_2 -driven limnic eruption and would be amplified, and probably dominated (depending on the magnitude of the volcanic eruption), by contribution from the limnic eruption. CO_2 -driven water volcanism has also been proposed to play a role in the resurfacing of the jovian moon Europa¹⁹.

CO_2 -driven limnic eruptions are similar to H_2O -driven violent volcanic eruptions in that they are both powered by exsolution of gas from a liquid, though there may be small differences regarding the role of buoyancy and turbulence. Buoyancy helps limnic eruptions (as the eruption is through fluid) but not violent volcanic eruptions. On the other hand, turbulence plays a greater role in reducing the erupting velocity in limnic eruptions (owing to the low viscosity of water and the instability of a foam) than in violent volcanic eruptions. Ignoring these differences, the violence of a gas-driven eruption is largely determined by the solubility coefficient (λ) and by the total decompression ratio (P_0/P), based on equation (3). Therefore, both CO_2 -water systems at room temperature and H_2O -magma systems at magmatic temperatures can power violent eruptions. On the other hand, CO_2 -magma (except for CO_2 -kimberlitic magma) systems may not be able to power violent eruptions owing to the low solubilities of CO_2 in rhyolitic to basaltic magma. □

Received 19 June; accepted 14 November 1995.

- Freeth, S. J. & Kay, R. L. F. *Nature* **325**, 104–105 (1987).
- Sigurdsson, H. et al. *J. Volcan. geotherm. Res.* **31**, 1–16 (1987).
- Kling, G. W. et al. *Science* **236**, 169–175 (1987).
- Sabroux, J. C. et al. *J. Volcan. geotherm. Res.* **42**, 381–384 (1990).
- Sigvaldason, G. E. J. *Volcan. geotherm. Res.* **39**, 97–107 (1989).
- Tazieff, H. J. *Volcan. geotherm. Res.* **39**, 109–116 (1989).
- Giggenbach, W. F. J. *Volcan. geotherm. Res.* **42**, 337–362 (1990).
- Kling, G. W., Tuttle, M. L. & Evans, W. C. J. *Volcan. geotherm. Res.* **39**, 151–165 (1989).
- Freeth, S. J. in *Natural Hazards in West and Central Africa* (eds Freeth, S. J., Ofegbu, C. O. & Onouha, K. M.) 63–82 (Vieweg, Braunschweig, 1992).
- Mader, H. M. et al. *Nature* **372**, 85–88 (1994).
- Zhang, Y., Sturtevant, B. & Stolper, E. M. in *IUGG XXI General Assembly B413* (abstr.) (1466, Boulder, CO, 1995).
- Kanari, S.-I. J. *Volcan. geotherm. Res.* **39**, 135–149 (1989).
- Tietze, K. in *Natural Hazards in West and Central Africa* (eds Freeth, S. J., Ofegbu, C. O. & Onouha, K. M.) 97–108 (Vieweg, Braunschweig, 1992).
- Evans, W. C. et al. *Geochem. J.* **28**, 139–162 (1994).
- Wilson, L. J. *Volcan. geotherm. Res.* **8**, 297–314 (1980).
- Cussler, E. L. *Diffusion: Mass Transfer in Fluid Systems* (Cambridge Univ. Press, 1984).
- Zhang, Y., Stolper, E. M. & Wasserburg, G. J. *Geochim. cosmochim. Acta* **55**, 441–456 (1991).
- Kling, G. W., Evans, W. C., Tuttle, M. L. & Tanyileke, G. *Nature* **368**, 405–406 (1994).
- Crawford, G. D. & Stevenson, D. J. *Icarus* **73**, 66–79 (1988).
- Wiebe, R. & Gaddy, V. L. *J. Am. chem. Soc.* **62**, 815–817 (1940).
- Weiss, R. F. *Mar. Chem.* **2**, 203–215 (1974).
- Dean, J. A. *Lange's Handbook of Chemistry* (McGraw-Hill, New York, 1985).
- Weast, R. C. *CRC Handbook of Chemistry and Physics* (CRC, Boca Raton, FL, 1983).

ACKNOWLEDGEMENTS. I thank W. C. Evans for reviewing the manuscript, and B. Sturtevant, S. J. Freeth, E. J. Essene, R. A. Lange, J. R. O'Neill, H. N. Pollack and B. H. Wilkinson for their constructive comments. I also thank C. Jones and J. R. O'Neill for helping create the word ‘ambioructic’. This work was supported by the US NSF.

An ancient wind-powered iron smelting technology in Sri Lanka

Gill Juleff

Samanalawewa Archaeological Project, Archaeological Department of Sri Lanka, Sir Marcus Fernando Mawatha, Colombo 7, Sri Lanka and University College London, Institute of Archaeology, 31–34 Gordon Square, London WC1H 0PY, UK

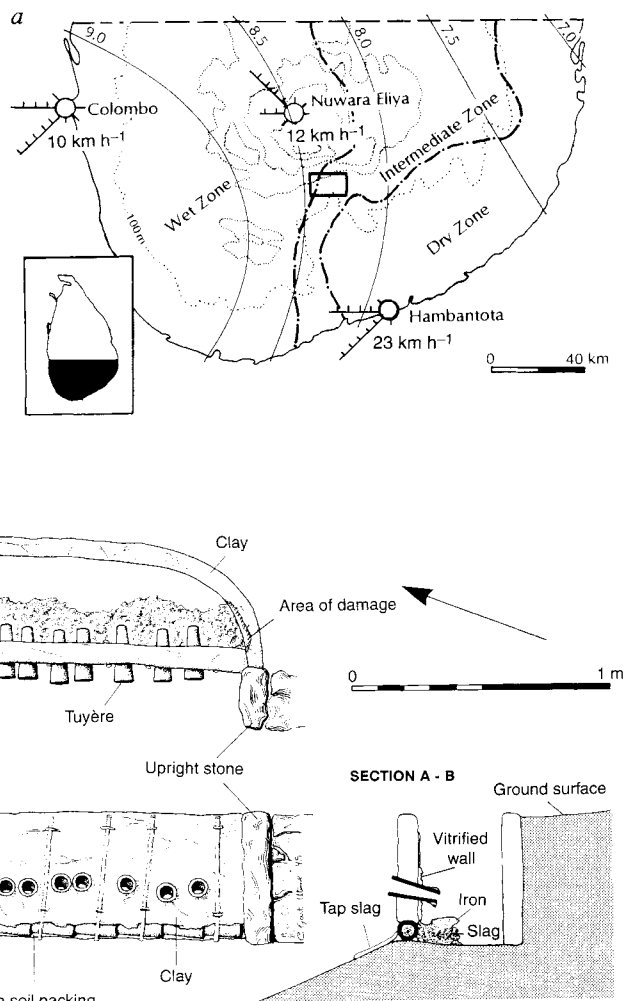
BEFORE the development of the blast furnace, iron smelting was achieved by ore reduction at temperatures below the melting point of the metal, forming an agglomerated 'bloom' of low-carbon iron and slag. The forced-draught (bellows-operated) shaft furnace known from archaeological studies is usually regarded as the pinnacle of this early smelting technology^{1–3}. Examples of natural-draught furnaces, in which gas buoyancy in a shaft of sufficient height induces a draught adequate to drive the smelting process⁴, are also known, but are generally regarded as disappointingly inefficient by comparison⁵. Here I report the discovery and excavation at Samanalawewa, Sri Lanka, of a previously unknown furnace type. The furnaces are all situated on the western margins of hills and ridges, where they are exposed to the strong monsoon winds. Field trials using replica furnaces confirm that this furnace type uses a wind-based air-supply principle that is distinct from either forced or natural draught, and show also that it is capable of producing high-carbon steel.

This technology sustained a major industry in this area during the first millennium AD, and may have contributed to South Asia's early pre-eminence in steel production.

Peninsular India, including Sri Lanka, is known for its rich ferrous archaeometallurgy; references, particularly those in classical and Islamic literature, to India as a source of high-quality steels have attracted considerable interest^{6–10}. Given the assumed limitations of bloomery smelting, such steels are regarded (with few exceptions) as necessarily the product of secondary refining processes. In South Asia, high-carbon crucible steels are well documented from the eleventh century AD onwards; a form of such a steel, known as *wootz*, was the raw material of mediaeval Indo-Islamic 'Damascus' swords^{8,11–15}. The investigation of the technology at Samanalawewa now indicates that comparable steels were produced directly, and in substantial quantities, in sophisticated 'frontal' smelting furnaces driven by wind-pressure.

Samanalawewa lies in hilly terrain between the lowland plains and central highlands of Sri Lanka (Fig. 1a). Its climate is dominated from June to September by desiccating winds associated with the southwest monsoon, which brings rain to the Wet Zone and *föhn* winds to the Intermediate and Dry zones¹⁶ (Fig. 1a). A plentiful supply of ore is ensured by numerous small hill-top deposits of secondary iron oxyhydroxides, common throughout southern Sri Lanka^{17,18}. Archaeological surveys, undertaken since 1988 as part of a hydro-electric scheme from which the area takes its name, have identified¹⁹ (within an area of 60 km²) 139 iron-working sites spanning 2,000 years. The technology described here is evidenced by 77 'west-facing' smelting sites, which form the largest component of the record. These are

FIG. 1 a Map of southern Sri Lanka showing the project area (within rectangle) and major climate divisions. Contour interval is 500 m unless marked. Also shown are July wind and pressure statistics for three meteorological stations, including percentage frequency of wind direction (in divisions of 10%), average wind speed (km h⁻¹) and pressure isobars (above 1,000 mbar)²⁰. Data collected from the study area over four monsoons (1990–92 and 1994) reveal a pattern of near-monodirectional, high-velocity winds of mean speed 31.5 km h⁻¹. Prevailing winds are west-northwest and west, although incident winds vary with local topography. b, Furnace used in smelting trials, reconstructed from archaeological data, comprising two elements; a semi-permanent rear wall and a temporary front wall. The rear wall is aligned north–south and curves westwards at each end to terminate with upright stones. The front wall, constructed between the two upright stones, is rebuilt with each smelt. From archaeological examples, furnace lengths (north–south axis) range from 1.2 to 2.3 m, while depth-in-plan (east–west axis) shows little variation, averaging 0.4 m. Furnace height, from base to rim of rear wall, is no more than 0.5 m. The supporting framework of sticks incorporated here into the front wall was the only feature for which there was no archaeological evidence. The front wall is designed to withstand the aggressive action of vitrification and slag formation during smelting, and to be broken at the end of the smelt. Greatest durability is imparted by the pre-fired, reused tuyères at the base where molten slag collects. These tuyères had already been used in the conventional manner, that is, as refractory conduits through which air entered the furnace. Multiple tuyères are generally equated with non-bellows furnaces. The re-used tuyères also act as a self-tapping mechanism, allowing slag to flow out of the furnace. Here the front wall was constructed abutting both rear wall and upright stones, whereas in archaeological examples it abuts the upright stones only. The stones thus serve as buffers between the two walls, helping to leave the rear wall free of damage in all but the extreme 'corners'.



all situated on the western margins of hill-tops and ridges, and the coincidence of the strongest incident winds during the monsoon with the locations of slag waste concentrations gave the first clue to wind utilisation.

Because metal products were rarely abandoned at their site of manufacture, slags are the principal indicators of furnace type and process for most research in extractive archaeometallurgy. In this case, the dominant forms are large, highly distinctive, sub-rectangular blocks of furnace slag. In these, slag has solidified against a horizontal line of reused tuyères; one tuyère telescoped into another to form the foundation of a straight furnace wall (Fig. 1b). To resolve the nature of the furnaces in which these slags formed, excavations were undertaken at one of the larger west-facing sites. The stratigraphic sequence of the site, covering $\sim 3,000 \text{ m}^2$ on a ridge aligned north-south, revealed regular, episodic activity devoted exclusively to smelting. Eight radiocarbon dates from charcoal place²⁰ the use of the site between the seventh and early eleventh centuries AD. Approximately 20% of the site volume was investigated, revealing 41 furnace structures, all precisely positioned along the western brow of the ridge and oriented to face the incident wind. Furnaces conform to the basic two-component design shown in Fig. 1b. Rear walls were well preserved *in situ* whereas front walls were reconstructed from associated scatters of slag and clay fragments from the dismantled structures. Vitrification patterns and damage to rear walls caused by slag removal indicated maximum temperatures along the front wall and in the two furnace 'corners'.

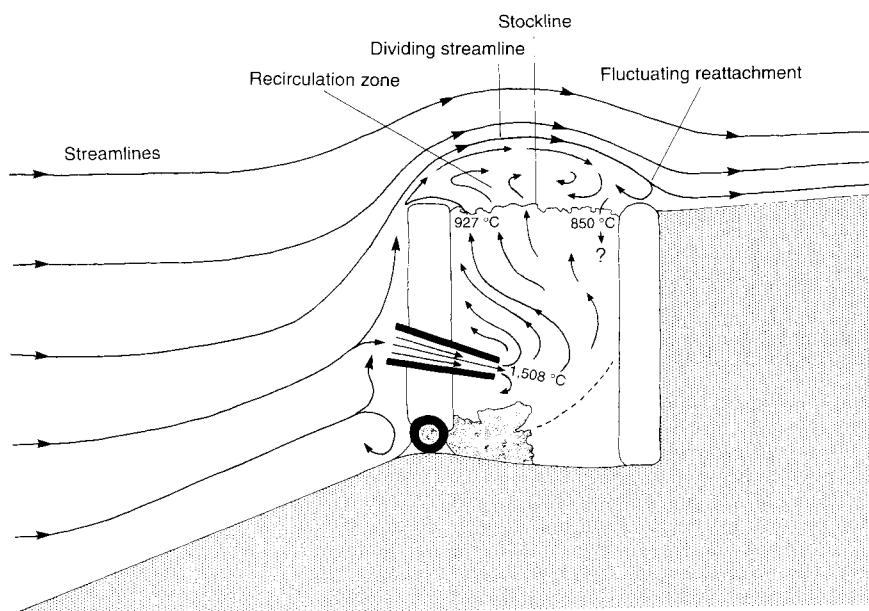
From survey data, and by extrapolation from quantitative sampling during excavation, the product output from the known sites is conservatively estimated at $\sim 3,500$ tonnes. Indications are that the industry, clearly successful, well organized and probably centrally controlled, was not restricted to Samanlawewa but extended into adjoining areas. Evidence from a single site within the area suggests the technology had roots in at least the third century BC (G.J., manuscript in preparation). Having reached a zenith in the ninth century, the industry disappears from the archaeological record in the eleventh century, probably as a result of incursions from South India which brought political

and demographic upheaval, and eventually the end of Sri Lanka's Dry Zone civilisation²¹.

'Wind-blown' iron smelting has not been reported with conviction from any part of the world and the possibility of such smelting had been largely discounted; the principal objection was based on the tacit assumption that the action of the wind would emulate that of bellows—that is, by blowing directly into the furnace (via tuyères)—and that gusting and calms, no matter how short, would produce an uncontrollable system liable to irreversible 'freezing'¹⁰. In addition, it was difficult to imagine that reducing conditions could be maintained in such an 'open' structure. To address these questions replication trials were carried out in July 1994, including three full smelts (trials 3, 4 and 5). Two new furnaces were constructed (Fig. 1b) at a west-facing site, ensuring that their orientation replicated that of archaeological furnaces at the site. Ore (79–87% Fe_2O_3) was collected from local deposits, and charcoal prepared from timber of a single species, *Syzygium spathulatum*; one of the three *Syzygiums* exploited preferentially for charcoal from the third century BC onwards. Wind speeds were recorded throughout, and during trial 5 furnace temperatures were measured using an optical pyrometer (Fig. 2). Progressive changes were made in ore/fuel ratios, from 1:1.5 to 1.5:1. In trials 4 and 5, the ore was roasted for 30 minutes over a charcoal fire before placing it (charging) into the furnace and, in trial 5, ore size was halved, from $\sim 30 \text{ mm}$.

Each smelt followed the same procedure, beginning with a two-hour preheat, burning charcoal only, followed by the gradual addition, over ~ 3.5 hours, of pre-weighed, alternately layered, ore/charcoal charges (total ore 81, 95 and 126 kg respectively). Tap slag formed readily, within an hour of the first ore charge, the time decreasing with successive smelts, and flowed through the line of re-used tuyères at the base of the front wall. After a further 1.5 hours charging charcoal only, the stockline was allowed to burn down from rim to half furnace height ($\sim 1.5 \text{ h}$). To end the smelt, the front wall was pushed inwards and the furnace contents (charcoal, slag, metal and the broken front wall) dragged westwards out of the furnace. Increasing ore/fuel ratios, ore roasting and reduction in ore size all contributed to enhanced reactivity

FIG. 2 Furnace cross-section showing observed and postulated airflow, and maximum combustion zone and furnace-top temperatures recorded during trial 5. Wind close to ground level, flowing upslope, reaches maximum acceleration at the crest of the hill where it encounters the furnace. As it is forced over the front wall, boundary layer separation occurs creating a recirculation 'bubble', capping the top of the furnace; low pressure within the bubble draws the dividing streamline downwards again until it reattaches at a point close to the rear wall. Within the bubble, which preserves low pressure, heat and reducing conditions, trapped combustion gases (clearly visible at night) flow along the north-south axis of the furnace perpendicular to the incident wind. The point of lowest pressure occurs towards the outer lip of the front wall and airflow within the furnace is towards this point, producing the smelting curtain or 'front' seen along the front wall. This 'front' remained dimensionally stable (at about one-third of furnace depth-in-plan) throughout the smelts and, although increasing wind speed intensified its incandescence, it did not increase its penetration into the charcoal bed. This observation demonstrates that there is little effect from wind blowing directly into the furnace, which in turn affords the system a measure of stability; its response to changes in wind speed is slow, negating the effects of gusting. In trial 5, a 1.5-hour period of low wind speeds ($< 15 \text{ km h}^{-1}$) caused only a gradual lowering of the mean combustion zone temperature from 1,454 to



1,336 °C. (The temperatures reached in trial 4 (mean wind speed 44 km h^{-1}), although not measured would have been in excess of those noted here.) Increased combustion at the front wall also results in a physical gradient allowing charcoal and ore to fall towards the front, rather than falling, unreacted, to the back of the furnace. The front wall, and in particular its outer lip, would be critical in creating the desired airflow and pressure regime and was probably modelled with greater precision in ancient times.

and yields with successive trials. The most successful charging regime was achieved in trial 5, yielding 17 kg metal despite low wind speeds (mean 22 km h^{-1}). The strongest winds prevailed during trial 4 (mean 44 km h^{-1}), giving the highest fuel consumption (2.2 kg per h per tuyère). Replication by the trials of characteristic features of the archaeological slags, furnace wall vitrification patterns and post-smelt debris arrays gives veracity to the interpretations drawn from the experiments.

The trials demonstrated visibly that high temperatures and smelting reactions are concentrated immediately inside the front wall (Fig. 2). From this, and observations of gas flow at the top of the furnace, it appears that the system is driven by wind flow over the front wall creating a low-pressure zone at the top of the wall. This results in a steep pressure gradient between the external tuyère mouth and the top of the furnace, which then pulls air through the tuyères. Fluid-mechanics calculations, based on wind velocities measured during trial 4 and using conservatively estimated parameters, predict a pressure drop of the order of

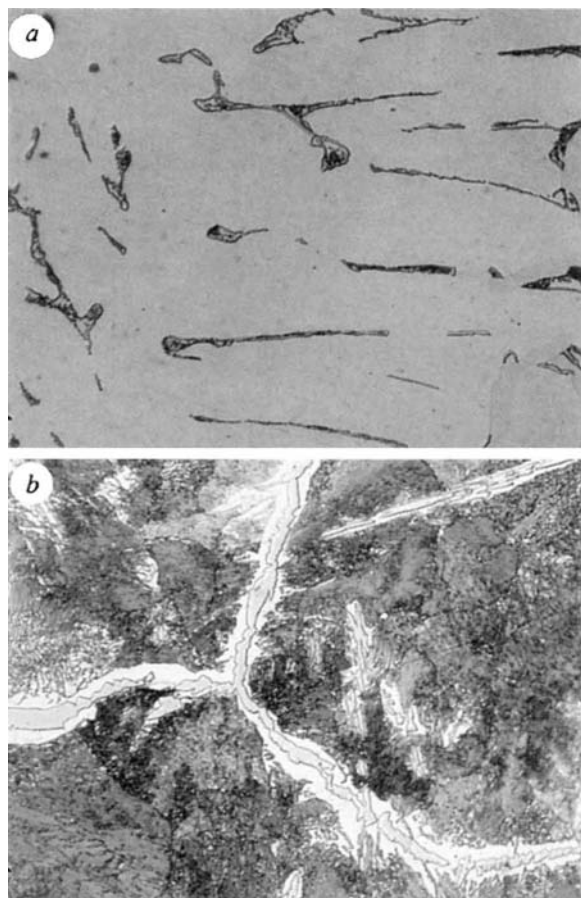


FIG. 3 a, Optical micrograph of low-carbon iron with estimated 0.1% C, consisting of ferrite with a small amount of pearlite. The sample shown was etched with nital; magnification $\times 710$. A 270-g fragment of similar material, from trial 4, was forged by local blacksmiths into a bar weighing 125 g, demonstrating that approximately half comprised entrapped slag. b, Optical micrograph of hypereutectoid steel typical of that from the top of the furnace slag, showing proeutectoid cementite (Fe_3C) on grain boundaries, in pearlite matrix. The pearlite is both lamellar (extreme left) and degenerate with some spheroidization (elsewhere). The estimated carbon content within this sample is 0.6 to $>1\%$, with the area shown being $>0.8\%$ C. The sample shown was etched with nital; magnification, $\times 710$ (M. L. Wayman, personal communication). The mechanism by which the metal lowest and longest in the furnace has absorbed carbon to such an extent has yet to be investigated, but parallels may be drawn with the mechanisms described for the *Tatara* process²⁴. Together, the products from the trials suggest successful, but not fully optimized, smelts. Given further experimentation, it is possible to envisage better consolidation of all the reduced metal, more uniform carburization, and complete metal/slag separation.

50 Nm^{-2} . In contrast, gas buoyancy, as in a natural-draught furnace, has little or no effect in this furnace, of only 0.35 m height (tuyère nozzle to furnace rim); the relevant calculation predicting a pressure drop of $< 3 \text{ Nm}^{-2}$ (D. J. Wilson, personal communication). The air supply mechanism described gives continuous aspiration, as compared with the intermittent pumping of bellows, which is recognized as decisive in achieving the high temperatures needed to promote carburization.

In characterizing this furnace, the term 'frontal', expressing both smelting zone and overall furnace geometry, is used to distinguish it from the axial symmetry of the shaft (and blast) furnace. This frontal configuration provides the largest smelting zone for a given furnace volume and allows variable capacity: once depth-in-plan (east–west axis) is optimized for the expected air supply regime, the furnace can be substantially enlarged by lengthening the structure. Two ancestral Sri Lankan furnace forms show progressive evolution of the 'frontal' concept, namely the third century BC Samanawewa example, and the first century BC to third century AD examples from Sigiriya²². Both have front walls that increase in length, over time, from 0.5 to 0.95 m, the depth-in-plan remaining at ~ 0.5 m. Similar frontal design is recognizable in other furnaces not driven by wind, for example, the natural-draught furnaces of Burma²³ and the bellows-operated *Tatara* of Japan²⁴, both used for iron smelting. The latter is one of the very few pre-modern processes capable of producing usable quantities of high-carbon steel. Whether historical connections exist between these Asian technologies deserves investigation. As the data stands, the west-facing furnaces seem to represent an independent development with no known parallels in India, the assumed cultural well-spring of Sri Lanka.

Metal products from the trials divide into two general types, stratified within the furnace: an upper layer of detached 'blooms', and a lower layer of well consolidated metal still adhering to the furnace slag (Fig. 1b). No metal was found embedded in archaeological slags, and incomplete metal/slag separation was a shortcoming in the trials. Metallographic analysis (Fig. 3) of the detached metal reveals generally low-carbon iron in forms typical of the varying stages in bloomery smelting²⁵. In contrast, and unexpectedly, the metal attached to the slag (comprising $\sim 50\%$ of the metal produced) is relatively slag-free and predominantly high in carbon, and can be classed as high-carbon steel. Carburization, although heterogeneous, is sufficiently extensive to indicate a sought-after product rather than fortuitous occurrence. This finding has important bearing on a ninth century Islamic reference^{8,14,26}, describing the use of *Sarandibi* (Sri Lankan) steels in sword-manufacture. Whereas previously it was assumed that this referred, by extrapolation from later accounts, to crucible steel, it now seems probable that the material in question was 'furnace steel'. Sri Lanka's well established Indian Ocean trade links makes distribution of products from the west-facing sites entirely feasible, and dated artefact finds from Sri Lanka²⁷ and the east coast of Africa²⁸ demonstrate that high-carbon steels were in circulation at the time. The data presented here provide the earliest dated field evidence from South Asia for the industrialized production of high-carbon steel. □

Received 18 August; accepted 5 December 1995.

1. Tylecote, R. F. *A History of Metallurgy* 2nd edn (The Institute of Materials, London, 1992).
2. Rostoker, W. & Bronson, B. *Pre-industrial Iron* (Archaeometallurgy Monogr. 1, Philadelphia, 1990).
3. Bamberger, M. in *Furnaces and Smelting Technology in Antiquity* (eds Craddock, P. T. & Hughes, M. J.) 151–158 (British Museum, London, 1985).
4. Rehder, J. E. *Archaeometallurgy* **2**, 47–58 (1987).
5. Killick, D. in *Recent Trends in Archaeometallurgical Research* (ed. Glumac, P.) 47–54 (Science and Archaeology Vol. 8, Univ. Pennsylvania, Philadelphia, 1991).
6. Hadfield, R. J. *Iron Steel Inst.* **85**, 134–186 (1912).
7. Schoff, W. H. *J. Am. Oriental Soc.* **35**, 224–239 (1915).
8. Bronson, B. *Archaeometallurgy* **1**, 13–51 (1986).
9. Chakrabarti, D. K. *The Early Use of Iron in India* (Oxford Univ. Press, Delhi 1992).
10. Craddock, P. T. *Early Metal Mining and Production* (Edinburgh Univ. Press, 1995).
11. Faraday, M. Q. *J. Sci. Lit. Arts* **7**, 319–330 (1819).
12. Zschokke, B. *Rev. Métall.* **21**, 635–669 (1924).
13. Maryon, H. *Stud. Conserv.* **5**, 25–37, 52–59 (1960).
14. Allan, J. W. *Persian Metal Technology 700–1300 AD* (Ithaca, London, 1979).
15. Wadsworth, J. & Sherby, O. D. *Prog. Mater. Sci.* **25**, 35–68 (1980).

16. Domrös, M. *The Agroclimate of Ceylon* (Franz Steiner, Wiesbaden, 1974).
17. Cooray, P. G. *An Introduction to the Geology of Sri Lanka* (National Museums of Sri Lanka, Colombo, 1984).
18. Herath, J. W. *Mineral Resources of Sri Lanka* (Geological Survey Dept, Colombo, 1980).
19. Juleff, G. *Ancient Ceylon* **9**, 75–107 (1990).
20. Malim, T. J. P., Juleff, G. & Shotliff, A. in *South Asian Archaeology 1995* (eds Allchin, F.R. & B.), (Oxford & IBH, New Delhi, in the press).
21. De Silva, K. M. *A History of Sri Lanka* (Hurst, California, 1981).
22. Forenius, S. & Solangarachchi, R. in *Further Studies in the Settlement Archaeology of the Sigiriya-Dambulla Region* (eds Bandaranayake, S. & Mogren, M.) 135–144 (Postgraduate Inst. Archaeology, Colombo, 1995).
23. Percy, J. *Metallurgy*, Vol. 2, 271–273 (Murray, London, 1864).
24. Rostoker, W., Bronson, B. & Dvorak, J. R. *Archeomaterials* **3**, 11–25 (1989).
25. Avery, D. H., van der Merwe, N. J. & Saitowitz, S. in *The Beginnings of the Use of Metals and Alloys* (ed. Maddin, R.) 261–282 (MIT Press, Cambridge, MA, 1988).
26. von Hammer Purgstall, J. *J. Asiatique* **5**, 66–80 (1854).
27. Cresswell, R. G. *J. hist. Metall. Soc.* **25**, 76–85 (1991).
28. Kusimba, C. M., Killick, D. & Cresswell, R. G. in *Society, Culture and Technology in Africa* (ed. Childs, S. T.) (Science and Archaeology Vol. 11, Univ. Pennsylvania, Philadelphia, in the press).
29. Somasekaram, T. (ed.) *National Atlas for Sri Lanka* (Survey Dept, Colombo, 1988).

ACKNOWLEDGEMENTS. I thank D. J. Wilson for the pressure calculations, M. L. Wayman for collaborating in the analysis of the metal products, T. Perry for assistance during the smelting trials, S. Davison and N. Seeley for comments on the manuscript, and C. Gait-Utme for the furnace illustrations. Special thanks also to S. U. Deraniyagala, P. B. Karunaratne, T. Malim and A. Shotliff for their contributions to the project. The Samanawewa project operates through the Archaeological Dept of Sri Lanka and was supported by British funding agencies and the Samanawewa Hydro-Electric Project.

Evolution of extreme specialization within a lineage of ectomycorrhizal epiparasites

K. W. Cullings*, T. M. Szaro & T. D. Bruns

Department of Environmental Science, Policy, and Management, University of California, Berkeley, California 94720 USA

MONOTROPE (Monotropoideae, Ericaceae) are achlorophyllous, epiparasitic plants that receive all of their fixed carbon from green plants through a common ectomycorrhizal association rather than by a direct parasitic connection (Fig. 1)^{1,2}. Using molecular identification methods we show that some monotropes are highly specific in their fungal associations and at least one species, *Pterospora andromedea*, is specialized on a single species group within the genus *Rhizopogon*. Phylogenetic analysis of the Monotropoideae shows that specialization has been derived through narrowing of fungal associations within the lineage containing *P. andromedea*. High specificity is contrary to past predictions for the Monotropoideae and for plant communities in general, raising many questions about the roles of mycorrhizal specificity in ecosystem function.

Ectomycorrhizae are mutualistic associations between a vascular plant root and a fungus; they are the dominant nutrient-gathering organs in most temperate forest ecosystems³. These mutualisms vary from general to specific, but when specificity occurs it is one-sided: fungal species may be specific to a single plant species or genus, but plants typically have dozens or even thousands of fungal associates^{4,5}. Plant specificity has been assumed to inhibit the ability of establishing seedlings to form new mycorrhizal associations^{6,7}, and because monotropes are entirely dependent on these associations, it has been suggested that they should be generalists⁶. This view was supported by a previous survey of monotrope associates that was based primarily on proximity of these plants to fungal sporocarps⁸.

To test the hypothesis that monotropes associate with many unrelated fungi, we used three different complementary polymerase chain reaction-based methods to identify fungal associates of four of the most common monotrope species: *Pterospora andromedea*, *Monotropa hypopithys*, *Monotropa uniflora*, and *Sarcodes sanguinea*. Results show a variety of fungal associates and levels of specificity within the monotropes (Table 1). *Sarcodes*

sanguinea appears to be a generalist that associates with at least three unrelated families of fungi. Although our sampling of *Monotropa* species was more limited, our results indicate that both may be specialists. *Monotropa hypopithys* was associated only with suilloid fungi, a monophyletic group that includes *Rhizopogon* and *Suillus*^{9,10}, whereas *M. uniflora* individuals collected from widely divergent habitats were all associated with fungi in the Russulaceae. *Pterospora andromedea* was found to be an extreme specialist. All 31 individuals, collected from a broad geographic region, were associated with the *Rhizopogon subcaerulescens* species group (Table 1). These results indicate a higher level of ectomycorrhizal specificity for *P. andromedea* than has been reported for any ectomycorrhizal plant. In comparison, *Alnus rubra*, which is considered to be exceptionally specialized, associates with 30 fungal species, many of which are only distantly related to each other¹¹.

We can reject the idea that specificity of *P. andromedea* is based on a lack of opportunity to associate with other fungi for the following reasons. (1) Ectomycorrhizal fungal diversity is very high within small pine monocultures; even adjacent root tips are frequently colonized by different fungi¹². (2) Suilloid fungi, though they produce numerous fruiting bodies in such forests, often account for a very low percentage of the colonized tree roots^{13,14}. (3) In the five locations where *P. andromedea* occurred with *M. hypopithys* or *S. sanguinea*, the latter two species were associated with different fungi (Table 1). (4) We analysed host tree roots present within the root ball of one *P. andromedea* and adjacent tree roots within 0.5 metres of three other *P. andromedea* plants and found that *R. subcaerulescens*, though present on the

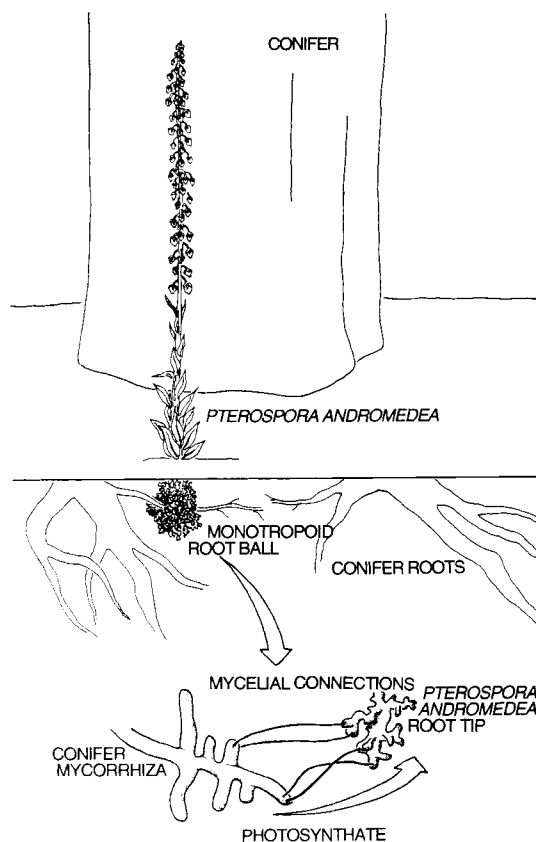


FIG. 1 Schematic summary of mycorrhizal epiparasitism in the Monotropoideae. Carbon fixed by the surrounding trees is transferred via a shared mycorrhizal fungus to the achlorophyllous monotrope: no direct connection between the two plants exists. The monotropes discussed in this paper have a compact root ball of mycorrhizae. The tree mycorrhizae, which are more diffuse in arrangement and more diverse in fungal associates, are often found within the root balls of the monotropes but are morphologically distinct.

*Present address Department of Biology, San Francisco State University, California 94132, USA

tree roots, was a minor component of the diverse fungal community present both within and near the rootballs. (5) At one site we collected fruit bodies of *Rhizopogon* species from the immediate vicinity of *P. andromeda*, and found that two ITS-RFLP patterns were represented in the three collections of *Rhizopogon* made. Neither of these patterns matched those from the *P. andromeda* samples even though one of these collections was made within 15 cm of a *P. andromeda* individual. Thus, *P. andromeda* clearly associates with a highly specific subset of the mycorrhizal fungal species present at a site.

Specificity is a common phenomenon in parasites, but detailed knowledge of how and why specificity evolves in parasitic lineages is often lacking^{15,16}. In this case we have a good estimate of the phylogeny of the monotropes from previous molecular phylogenetic analysis; this shows that the genus *Monotropa* is clearly polyphyletic and that the closest relatives of the monotropes are likely to be chlorophyllous members of the genus *Arctostaphylos*¹⁷ (Arbutoideae, Ericaceae). Overlaying the pattern of mycorrhizal associations onto the tree reveals that specificity is derived by a narrowing of mycorrhizal association from generalist, to suilloid specialist, and ultimately to specialization on *R. subcaerulescens* (Fig. 3). One could suggest that this resulted from fungal hosts evolving defences against the monotropes or by a gradual loss of host opportunity through ecological narrowing; in these scenarios, suilloids are the only fungi without an effective defense or the only ones common within the narrowed habitat. However, the rarity and diminutive size of these plants argues against them exerting significant selective pressure on the wide range of other common

fungi. The frequency of other mycorrhizal fungi in these habitats also shows that monotrope specificity is not based on a lack of non-suilloid fungal associates.

We argue that specificity resulted from a one-sided selection on these parasitic plants to increase their fitness by optimizing carbon allocation, rather than from co-evolution in the sense of reciprocal selection¹⁸. Suilloid fungi are among the most host-specialized ectomycorrhizal associates of the Pinaceae, many being restricted to single genera or species groups^{19,20}. Furthermore, although *Suillus* and *Rhizopogon* species can form mycorrhizae with non-host plants, these associations are usually aberrant in form, appear to have reduced function, and can result in preferential nutrient transfers from the 'non-host' to the specialized pineaceous host^{20,21}. Thus, it is likely that physiological accommodation by the monotropes would be necessary in order for them to benefit fully from a suilloid association. But what is the benefit of this accommodation if the price is the loss of opportunity to form associations with other common fungi? The same question could be applied to the suilloids: what have these fungi gained by specializing on the Pinaceae that outweighs the loss of opportunity afforded by a generalist strategy? One obvious hypothesis is that the suilloids, through specialization, have become more efficient at obtaining photosynthate from their hosts, and the most widespread monotropes in these pineaceous ecosystems have in turn specialized on the most efficient conduit of fixed carbon.

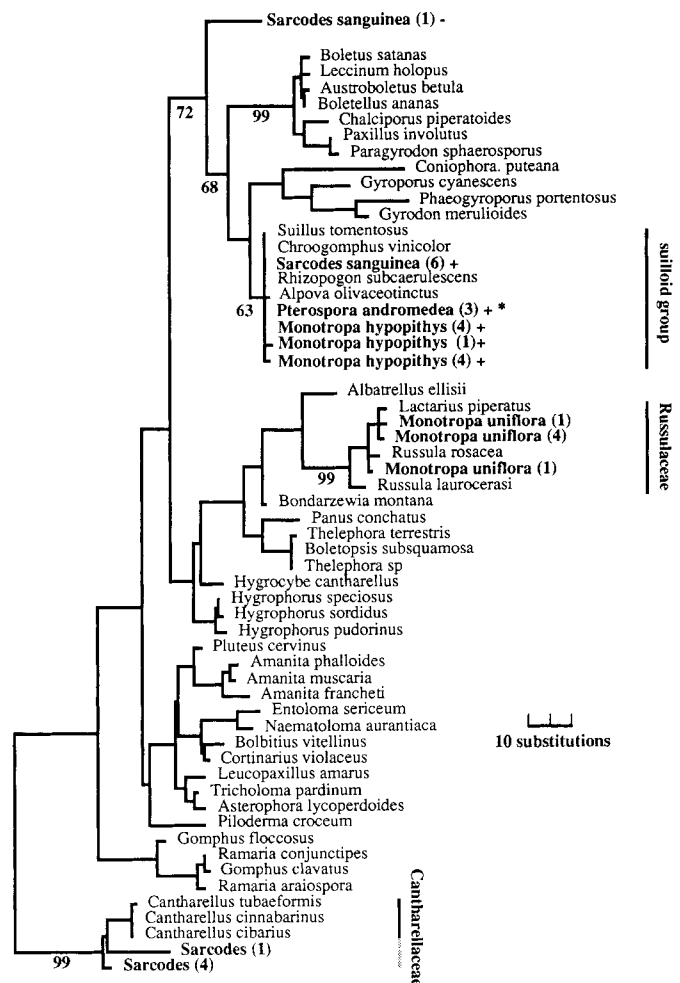
Regardless of whether this hypothesis is correct, these results demonstrate that an extreme in specialization on the plant side of ectomycorrhizae does exist, that it exists precisely where it was not

TABLE 1 Molecular identifications of fungal associates

Monotrope taxa	Number of plants	Fungi identified	Methods	Monotrope taxa	Number of plants	Fungi identified	Methods
<i>P. andromeda</i> (31)	3	suilloid gr.	Mt-LrRNA seq.	<i>M. uniflora</i> (6)	6	Russulaceae	Mt-LrRNA seq.
	30	suilloid gr.	TSOP	<i>S. sanguinea</i> (12)	6	suilloid gr.	TSOP and Mt-LrRNA seq.
	30	<i>Rhizopogon</i> or close relatives	TSOP				
	31	<i>R. subcaerulescens</i> gr.	ITS-RFLP and ITS seq.				
<i>Monotropa hypopithys</i> (9)	9	suilloid gr.	Mt-LrRNA seq.		4	<i>Rhizopogon</i> or close relatives	TSOP
	2	<i>Rhizopogon</i> or close relatives	TSOP		2	<i>Suillus</i>	TSOP
	1	<i>Suillus</i>	TSOP		5	Cantharellaceae	Mt-LrRNA seq.
	6	suilloid gr. - genus not clear	TSOP		1	non-suilloid unknown fungi	Mt-LrRNA seq., TSOP

Identification methods. Mt-LrRNA seq was based on sequence determination of the ML5/ML6 portion of the mitochondrial large subunit gene followed by phylogenetic placement of the unknown sequences (Fig. 2); Results of this method allowed us to identify most of the fungi to the level of family or subfamily. TSOP is based on taxon-specific oligonucleotide probing of the adjacent ML7/ML8 region of the same gene²². The five probes used specifically identify the Suilloid group (probe US1) as whole, which is a monophyletic lineage consisting of *Rhizopogon*, *Suillus*, *Truncocolumella*, *Alpova*, *Melanogaster* and the Gomphidiaceae, and the following taxonomic groups within it: portions of the Gomphidiaceae (probes G1 and G2), *Rhizopogon* and close hypogeous relatives (probe R1), and *Suillus* (probe S1). Identification to the suilloid group only, as in some *Monotropa hypopithys* samples, was based on strong hybridizations to the US1 probe and an absence of hybridization to the more specific probes. *Suillus tomentosus*, the most common species of *Suillus* in the Yellowstone area, would produce such a result, as it does not hybridize to the S1 probe²². Sequences, exact specificities and hybridization conditions have been described previously²². Fungal symbionts of *P. andromeda* were further identified to species by analysis of the internal transcribed spacer region (ITS), which is a nuclear encoded portion of the rDNA repeat unit²³. We used restriction fragment length polymorphisms (RFLPs) of the ITS region to screen all *P. andromeda* root samples and to compare these samples to identified fungal samples collected from the same sites. The enzymes *AluI*, *Hinf I*, and *MboI* were used as described elsewhere²³. A total of six different RFLP patterns was found. One of these was an exact match to that derived from a *R. subcaerulescens* collection from the Oregon site; the others appeared to differ only slightly from this pattern. We then sequenced the ITS region from representatives of each different ITS-RFLP pattern and found that sequence differences among the patterns varied from 0.3 to 2%. Because this level of variation is consistent with that found within other basidiomycete species or species groups²⁴⁻²⁶ we conclude that all *P. andromeda* root samples examined were associated with a member of the *R. subcaerulescens* species group. Note that this result is also consistent with the TSOP and Mt-LrRNA tests. **Sampling.** Total numbers of plant individuals sampled are given in parentheses following the monotrope species; numbers from this sample that were tested and identified with each diagnostic method are given in column two. Except for *M. hypopithys*, which was sampled only from Yellowstone Park, samples were taken from diverse habitats representing the full spectrum of environmental preferences of each species, and 1–3 individuals were taken from each site sampled. Collection locations follow: *P. andromeda*, California Sierra Nevada Mountains (mixed conifer, 10 sites in 3 counties); Oregon (*Pseudotsuga* dominated forest, 1 site), Yellowstone National park, Wyoming (*Pinus contorta* dominated forest, 6 sites, one of which had soils modified by geyser activity); *M. hypopithys*, Yellowstone National Park, Wyoming (*Pinus contorta* forest, 6 sites); *M. uniflora*, near Corvallis Oregon (mixed conifer/deciduous forest, 2 sites), Washington/Oregon border (sand dune ecosystem, 2 sites) and Minnesota (mixed conifer/hardwood, 2 sites); *S. sanguinea*, central Sierra Nevada mountains of California (mixed conifer forest, 6 sites in 3 counties).

FIG. 2. Phylogenetic placement of ectomycorrhizal fungi associated with the Monotropoideae based on sequence analysis of a 350-base-pair region of the mitochondrial *LrRNA* gene. Known fungi are shown in light type; unknown fungi are shown in bold and named by the monotrope species with which they were associated. The small size and conserved nature of the sequence limits the phylogenetic resolution of this tree. However, placement of the unknowns within tight terminal clusters such as the Russulaceae and the Cantharellaceae are strongly supported by bootstrap analysis; numbers at key branches are based on 500 bootstrap replicates. In the case of the suilloid group, unknown sequences differ from known suilloid sequences by 0 to 0.5%, and are confirmed as suilloid (+) within the cluster and as non-suilloid (–) outside it by oligonucleotide probes targeted at other regions (Table 1). *Pterospira andromedea* samples are further confirmed as suilloid by ITS RFLP and sequence analysis that place them within the *R. subcaerulescens* species group (*) (Table 1). The tree shown is a subset of a larger neighbour-joining analysis of Kimura two parameter distances for 103 taxa of identified basidiomycetes representing 42 genera in 12 families. The pruned taxa are additional members of clades already well represented in the above tree and their presence or absence does not affect the placement of the unknown sequences. Gaps introduced for alignment purposes were treated as missing data. Bootstrapping, Kimura distances, and tree construction were based on the SEQBOOT, DNADIST and NEIGHBOR programs of the PHYLIP 3.5c package²⁷ compiled for a Sun Sparc station. Parsimony analysis based on PAUP 3.1.1 (ref. 28) on a Macintosh Quadra yielded the same placement of unknowns within the groups indicated when the subset of taxa shown was used; the full data set, however was too large for parsimony analysis with our computers. The root of the tree shown (the Cantharellaceae) was arbitrarily chosen for convenience of display, but it is near the midpoint by parsimony criteria. Horizontal distances correspond to the minimum number of inferred substitutions (that is, parsimony criteria); vertical distance is arbitrary. Details concerning this data set will be reported in more detail elsewhere (T. D. B. *et al.*, manuscript in preparation)



expected^{6,8}, and that it occurs with fungal taxa that were thought to be nearly exclusive specialists of members of the Pinaceae. Furthermore, it means that key species of fungi are essential for the existence of these plants. This point is in contrast to the commonly espoused view that microbial species are functionally redundant.

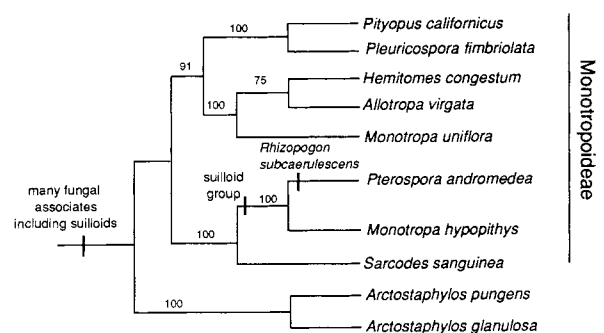
Although the Monotropoideae are rare plants and their specialization may represent an extreme, it is clearly an extreme in a continuum that has close evolutionary connections with *Arctostaphylos*¹⁷ (Fig. 3), one of the most common genera of understory shrubs in pinaceous ecosystems of western North America. Thus, specialization by these epiparasites demands

that we re-evaluate our view of the evolution and role of specificity within the larger realm of ectomycorrhizal community structure and ecosystem function. □

Received 1 June; accepted 13 October 1995.

1. Björkman, E. *Physiol. Pl.* **13**, 308–327 (1960).
2. Leake, J. R. *New Phytol.* **127**, 171–216 (1994).
3. Read, D. J. *Experientia* **47**, 376–391 (1991).
4. Malloch, D. W., Pirozynski, K. A. & Raven, P. H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2113–2118 (1980).
5. Borowicz, V. A. & Juliano, S. A. *Evol. Ecol.* **5**, 385–392 (1991).
6. Molina, R., Massicotte, H. & Trappe, J. M. in *Mycorrhizal Functioning: An Integrative Plant-Fungal Process* (ed. Allen, M. F.) 357–423 (Chapman and Hall, New York, 1992).

FIG. 3 Evolution of extreme specialization within a clade of the Monotropoideae. Cladogram is taken from a previous study and is based on parsimony analysis of partial 28S rRNA gene sequences¹⁷. Numbers at nodes are bootstrap indices of support based on a sample of a 1,000 replicates. Bars represent inferred location of changes in host association within the *P. andromedea*, *M. hypophitys*, and *S. sanguinea* clade. *Arctostaphylos* spp. of the Arbutioideae (Ericaceae), the subfamily most closely related to the Monotropoideae¹⁷, are common in the understory of coniferous forests, and are known to be generalists that associate with many fungi including *Rhizopogon* species^{19,29}. *Sarcodes sanguinea* is a generalist with a similar, though possibly narrower range of associates that frequently includes suilloid fungi. *Monotropia hypophitys*, the closest relative of *P. andromedea*, appears to be restricted to associations with suilloid fungi, while *P. andromedea* associates exclusively with the *R. subcaerulescens* species group across its western geographic range. The clade containing *P. andromedea*, *M. hypophitys*, and *S. sanguinea* is strongly supported as a monophyletic group (bootstrap > 99%). The association of *M. uniflora* with the Russulaceae is not indicated with a bar because we do not know the fungal associates of other members of its clade. Note that the genus *Monotropia* as represented by this tree is polyphyletic. This conclusion is also corroborated by biochemical and



morphological character sets¹⁷. Furthermore, the polyphyletic nature of this genus is supported by the finding that the *M. hypophitys* lineage associate specifically with suilloid fungi while *M. uniflora* associates only with fungi in the Russulaceae.

7. Van der Plank, J. E. *Genetic and Molecular Basis of Plant Pathogenesis* (Springer, New York, 1978).
8. Castellano, M. A. & Trappe, J. M. *Mycologia* **77**, 499–502 (1985).
9. Bruns, T. D., Fogel, R., White, T. J. & Palmer, J. D. *Nature* **339**, 140–142 (1989).
10. Bruns, T. D. & Szaro, T. M. *Molec. Biol. Evol.* **9**, 836–855 (1992).
11. Miller, S. L., Koo, C. D. & Molina, R. *Can. J. Bot.* **69**, 516–532 (1991).
12. Bruns, T. D. *Pl. Soil* **170**, 63–73 (1995).
13. Natarajan, K., Mohan, V. & Ingelby, K. *Soil Biol. Biochem.* **24**, 279–280 (1992).
14. Danielson, R. M. *Can. J. Bot.* **62**, 932–939 (1984).
15. Thompson, J. N. *The Coevolutionary Process* 1–376 (University of Chicago Press, 1994).
16. Price, P. W. *Evolutionary Biology of Parasites* 1–237 (Princeton University Press, Princeton, NJ, 1980).
17. Cullings, K. J. *Evol. Biol.* **7**, 501–516 (1992).
18. Janzen, D. H. *Evolution* **34**, 611–612 (1980).
19. Molina, R. & Trappe, J. M. *New Phytol.* **126**, 653–675 (1994).
20. Molina, R. & Trappe, J. M. *Forest Sci.* **28**, 423–458 (1982).
21. Finlay, R. D. *New Phytol.* **112**, 185–192 (1989).
22. Bruns, T. D. & Gardes, M. *Molec. Ecol.* **2**, 233–242 (1993).
23. Gardes, M. & Bruns, T. D. *Molec. Ecol.* **2**, 113–118 (1993).
24. Baura, G., Szaro, T. M. & Bruns, T. B. *Mycologia* **84**, 592–597 (1992).
25. Vilgaly, R. & Sun, B. L. *Proc. natn. Acad. Sci. U.S.A.* **91**, 4599–4603 (1994).
26. Kasuga, T., Woods, C., Woodward, S. & Mitchelson, K. *Curr. Genet.* **24**, 433–436 (1993).
27. Felsenstein, J. PHILIP (Computer program, Univ. Washington, Seattle, 1995).
28. Swofford, D. L. PAUP (Computer program, Smithsonian Institution, Washington, DC, 1993).
29. Molina, R. & Trappe, J. M. *New Phytol.* **90**, 495–509 (1982).

ACKNOWLEDGEMENTS. We thank Mike Castellano, Dan Luoma, David McLaughlin, Jane Smith, and Jim Steele for assistance in obtaining field collections; Monique Gardes for helping on the identification of the tree mycorrhizae; the staff at Yellowstone, particularly Bob Lindstrom, and Sequoia National Parks for permits and logistical support; Amy David for artwork; Bob Patterson, Peter Price, Randy Molina, Jim Trappe, Jane Smith, and Detlev Vogler for constructive comments on the manuscript; and the Hardman Foundation and the NSF for financial support.

A member of the *KNOTTED* class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*

Jeff A. Long*, Erich I. Moan†, June I. Medford‡ & M. Kathryn Barton†§

† Department of Genetics and * Program in Cellular and Molecular Biology, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

‡ Department of Biology, 506 Wartik Laboratory, Pennsylvania State University, College Park, Pennsylvania 16802, USA

THE *KNOTTED* class of plant genes encodes homeodomain proteins^{1,2}. These genes have been found in all plant species where they have been sought^{3–5} and, where examined, show expression patterns that suggest they play an important role in shoot meristem function^{5–7}. Until now, all mutant phenotypes associated with these genes have been due to gain-of-function mutations^{4,5,8,9}, making it difficult to deduce their wild-type function. Here we present evidence that the *Arabidopsis SHOOT-MERISTEMLESS (STM)* gene, required for shoot apical meristem formation during embryogenesis, encodes a class I *KNOTTED*-like protein. We also describe the expression pattern of this gene in the wild-type plant. To our knowledge, *STM* is the first gene shown to mark a specific pattern element in the developing plant embryo both phenotypically and molecularly.

The angiosperm shoot apical meristem (SAM) forms during embryogenesis and acts to generate leaves, stem and floral organs throughout the lifetime of the plant. *Arabidopsis* seedlings homozygous for recessive mutations induced by ethylmethane sulphonate (EMS) at the *SHOOTMERISTEMLESS (STM)* locus fail to develop a SAM during embryogenesis¹⁰ (Fig. 1). The converse phenotype, the development of many, ectopic SAMs, has been observed in tobacco plants that constitutively express the wild-type maize *KNOTTED1* gene¹¹). Because gain-of-function and loss-of-function mutations in regulatory genes often result in opposite phenotypes, we hypothesized that the *stm* mutations

were the result of loss-of-function mutations in an *Arabidopsis KNOTTED*-like gene. An *Arabidopsis KNOTTED* homologue, the meriHB1 cDNA, has been isolated (C. Granger *et al.*, manuscript submitted) that is located near *STM* on chromosome I. No recombination was found between *stm-1* and a restriction-fragment length polymorphism (RFLP) detected by meriHB1 (0/74 chromosomes tested), indicating that *stm-1* is within 5 centimorgans (cM) of meriHB1 (95% confidence interval).

The genomic DNA sequence corresponding to meriHB1 was determined for two mutant alleles of *STM*; both *stm-1* and *stm-3* harbour mutations predicted to reduce gene function dramatically (Fig. 2). The *stm-1* allele carries a nonsense mutation upstream of the homeodomain. The *stm-3* allele carries a mutation at the 3' splice acceptor site of the first intron. Similar mutations greatly reduce the use of this splice acceptor site in a variety of species^{12–14}. Based on the close linkage of *STM* to the meriHB1 gene and the existence of deleterious mutations in the *stm-1* and *stm-3* alleles, we conclude that the *STM* locus encodes an *Arabidopsis KNOTTED* homologue and thus is likely to promote SAM formation through transcriptional regulation.

The predicted *STM* protein is 382 amino acids long and is most similar to the *SBH1* gene from soybean³ (67% identity). It is roughly equally similar to maize *KN1*¹ (47% identity), *Arabidopsis KNAT1*⁵ (47% identity) and *Arabidopsis KNAT2*⁵ (44% identity). In the homeodomain, *STM* is again most similar to the soybean homologue (95% identity) and somewhat less similar to *KN1*, *KNAT1* and *KNAT2* (92, 89 and 83% identity, respectively). *STM* is a class I *KNOTTED*-like gene based both on its high degree of similarity to *KNOTTED* in the homeodomain and on its expression pattern in the plant (see below). Three blocks of conserved

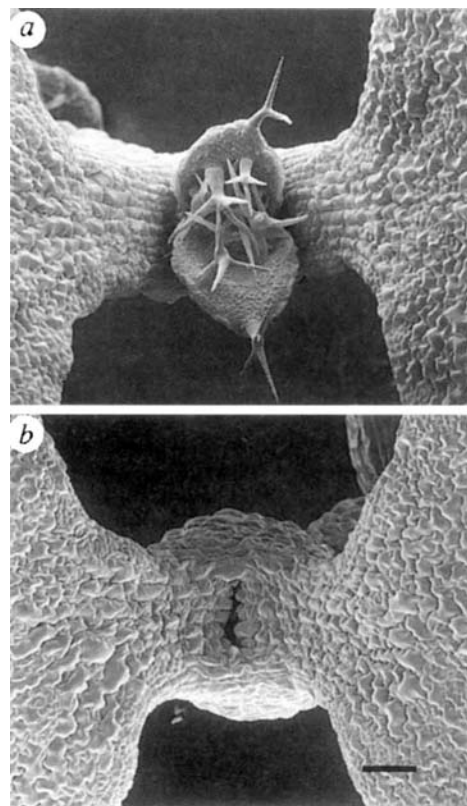


FIG. 1 Scanning electron micrographs of ~1-week-old wild-type (a) and homozygous *stm-1* (b) seedlings. Wild-type seedlings have a functional SAM as determined by the presence of two leaves at the bases of the cotyledons. Seedlings homozygous for the strong *stm-1* allele do not develop a shoot apical meristem. Scale bar, 100 μ m. Seedlings were prepared and visualized by scanning electron microscopy as described in ref. 17.

§ To whom correspondence should be addressed.

FIG. 2 Coding sequence of the wild-type *STM* gene (WS ecotype) aligned with coding sequences of soybean SBH1³, maize *KNOTTED1*, and *Arabidopsis KAT1* and *KAT2* genes⁵. Arrows indicate the positions of introns in *STM*. The extent of the homeodomain is indicated by the solid line. Introns are present at similar (introns 1 and 2) or identical (intron 3) positions in *KAT1*, *KAT2* and *KNOTTED1*. The *STM* gene lacks an intron found in the other homologues (between introns 1 and 2). The five-pointed star indicates the position of the glutamine to ochre stop codon change (CAA to TAA) in *stm-1*. The asterisk indicates the position of the mutation found in *stm-3*. This mutation changes the 3' splice acceptor site from AG GT to AA GT.

METHODS. The meriHB1 cDNA extends from aminoacid position 81 to the 3' end of the transcript and includes a poly(A)⁺ tail (C. Granger et al., submitted). A probe made from meriHB1 detects a codominant RFLP between *Niederzetz* and *Landsberg erecta* ecotypes in *EcoRI*-digested genomic DNA. To map this RFLP relative to *stm-1*, *Landsberg erecta* plants heterozygous for *stm-1* were crossed by *Niederzetz* and recombination was assessed in the F₂ self-progeny of F₁ hybrids (F₃ seedlings were used both to determine the genotype of each F₂ plant and as a source of DNA.). For the wild-type sequence determination, the meriHB1 cDNA was used to isolate a genomic clone (WS ecotype). This genomic clone (12 kb) was subcloned into the pMON754 vector (Monsanto). The sequence of 3.2 kb of this genomic DNA was determined using Sequenase 2 (USBiochemical) and synthesized oligonucleotides (Operon Technologies). The predicted *STM* coding sequence 5' of the cDNA sequence was deduced from the genomic DNA sequence (C. Granger et al., submitted). This 5' sequence results in a predicted amino-acid sequence that is similar to the most closely related

STM	ME--S-----	-----GSNSTSCPMFAGDNDGPMCPMMMPIMTSH	37
SBH1	MEGGSS-----	-----SSNGTSYLLAFGENSSG-LCPMTMM--PLVTS	37
KN1	ME--EITQHF-----	-----VGASSHGHCQHCHHHHHHPWASSL	37
KNAT1	ME--EYQHDNSTTPQVRSFLYSPISSSNKNDNTSDTNNNNNNN--SSNY		47
KNAT2	MD--R-----		4
STM	Q--HHGHDHQHQOEHGDAYQ--SHH--CQSS-----	SLFLQSLAPPGTK--NKVASSSSPSS--CAPAYS--	96
SBH1	HAGHHFIPNSNNNNVNTCLFIPNCSNS--TGTP-----	SIMLHN--NNNN--NKTDDDDNNNN--TGLGY--	97
KN1	-----SAVVA--PLP--PQPP-----	SAGLPL--TLNT--VAATGNSGGSGNPVLQLANGG--	81
KNAT1	-----GPCYN--NTN--NNHHHQHMLFPHMSLLPQ--	TTENCFRSDHQDQNNNNNPVSKSEASSSRIN	105
KNAT2	-----MCGFR--STEDYSEKA-----	TLMPFS--DYQS--LICSTTGDNR--LFGSDE--	45
		↓	
STM	----LMEI--HNEIVAGGINPCSSFSASVAKIMAHPHYHRLLAAYVNCQVGAPEVVARLEECSSAA--AAAAAMGPTG		172
SBH1	----FMESDHHHHHGNNNNGSSSSSSSAVKAKIMAHPHYHRLLAAYVNCQVGAPEVVARLEECSSAAATAGGDAAGSS		177
KN1	----LLDA--CVKAKEPSSSSP--YAGD--VEAIKAKIISHPHYSLTAYLE--NKVGAPEVVARLEECSSAA--ARQRTLQGL		154
KNAT1	HYSLMLRA--IHTYQEAANNNDNVSD--VEAMKAKIAHPHYSTLLQAYLDCQKIGAPDVDRITAAQDFFE--ARQRTSTPS		185
KNAT2	----LATA--LSSELLPRIKKAEDNFS--LSVTKSKIASHPLYRLLQTYIDCQVGAPEMETACILEEIQRENH--VYKRDVAPLS		120
		↓	
STM	CLGE--DPGLDQFMEAYCEMLVKYBQELSKPFKEAMVFLQVCEQPKSL-----	SLSSPSSSFGYETAIDRNNNGSSSEEDVDMN	250
SBH1	CIGE--DPALDQFMEAYCEMLVKYBQELSKPLKEAMFLQVCEQPKNL-----	TISS--SDFA--SNEGGRNGSSSEEDVDMN	250
KN1	AAAT--EPELDQFMEAYHEMLVKFREELTRPLQEAMFMRVRESQLNSL-----	SISG--RS--L--RNILSSGSSSEEDQEGSGG	226
KNAT1	SASSRDPELDQFMEAYCEMLVKYBQELTRPLQEAMFIRIESQLNSLCQSPHILHN--PD--G--KSDNMSSSDEEQEANS		264
KNAT2	CFGA--DPELDEFMETYCDILVKVYKTDRLRPFDEATFTINKENQJLQ-----	CTGP--AS--A--TALSDGAVSSDEELR	192
	↓		
STM	N---EFVDPQ---AEDRELKQQLLRKYSGYLSLKQEFMKKRGKGLPKPEARQQLLDWNSRHYKWPYPSEQKLAESTGL		326
SBH1	N---M-IDPQ---AEDRLKQQLLRKYSGYLSLKQEFMKKRGKGLPKPEARQQLLDWNSRHYKWPYPSEQKLAESTGL		325
KN1	ETELPE--VDPAH---GVDQELKHLLKRYSGYLSLKQELSKKGGKGLPKPEARQQLLDWNSRHYKWPYPSETQKVAESTGL		305
KNAT1	ETELPE--IDPR---AEDRELKQQLLRKYSGYLSLKQELSKKGGKGLPKPEARQQLLDWNSRHYKWPYPSEQKVAESTGL		343
KNAT2	D---D-IAADDSQQRSDNRDLKQQLLRKFGSHISSLKLEFSKKKGGKGLPREARQALLDWNVHNKWPYPTEGDKISLAESTGL		272
		↓	
STM	DQKQINWFIQNRKRHWKPSSEDMQFVMD---AT--HPHHYFMD-----	NVLNPFPMHISSTML--	383
SBH1	DQKQINWFIQNRKRHWKPSSEDMQFVMD--PS--HP-HYFMD-----	NVLGNPFPM-DLSHPM--L	379
KN1	DLKQINWFIQNRKRHWKPSSEMHLMMDGYHTT--NA-FYMDGHPINDGGLYRLG		358
KNAT1	DQKQINWFIQNRKRHWKPSSEDMQFVMDGLQHPHAA--LYMDGHPMDGP--YRLG-----	P-	399
KNAT2	DQKQINWFIQNRKRHWKPSSEMTMD--DS--NE-TFPT-----		311

homologue, SBH1. There are stop codons in all three reading frames upstream of the predicted initiating methionine. Intron positions were deduced by comparing the genomic sequence to the cDNA sequence. To determine the sequence of mutant alleles, genomic DNA from mutants was amplified by the polymerase chain reaction using either *Taq* polymerase (Promega, *stm-1*) or *Pfu* (Stratagene, *stm-3*). The DNA sequence of the resulting PCR products was either determined directly (*stm-1*; Epicentre Sequitherm) or cloned into pUC18 and the sequence of two independent clones subsequently determined (*stm-3*). Because both *stm-1* and *stm-3* were isolated in the *Landsberg erecta* ecotype, wild-type *Landsberg erecta* genomic DNA was amplified, sequenced and used as a comparison. Sequences were aligned using the Jotun Hein method (DNASTAR software).

sequences are found upstream of the homeodomain. Two fall in a region (181–220) predicted to form an amphipathic helix that may play a role in protein–protein interaction. The third, the ELK domain, is just upstream of the homeodomain; acidic residues upstream of this have been suggested to act as a transcriptional activation domain².

In homozygous *stm-1* embryos, no SAM is initiated¹⁰. In wild-type torpedo-stage embryos, cells at the presumptive apex adopt division patterns characteristic of the SAM and become organized into tunica and corpus layers. In contrast, SAM progenitor cells in *stm-1* mutants fail to adopt these division patterns and appear not to divide past this stage¹⁰. To characterize the expression of the *STM* gene in wild-type development, we performed *in situ* hybridization experiments. We find that *STM* messenger RNA appears before SAM initiation in the wild-type embryo and accumulates only in those cells predicted to form the embryonic SAM.

Dermatogen-stage embryos lack detectable *STM* expression (Fig. 3a). *STM* mRNA is first apparent in early to mid-globular-stage embryos (32–64-cell stage) where it is found in one or two cells (Fig. 3b). In late globular-stage embryos, *STM* mRNA appears in a stripe across the top half of the globular embryo (Fig. 3c). These cells are not clonally related¹⁵. As the cotyledon primordia grow out, the embryo enters the heart stage. *STM* expression is restricted to the notch between the cotyledons in the heart (Fig. 3d), torpedo (Fig. 3e) and mature-stage embryos (Fig. 3f). Control experiments using the sense strand of *STM* or the antisense strand of the *UBQY* gene result in either no signal or a uniform signal throughout the embryo, respectively (Fig. 3g, h). There is remarkable agreement between cells that accumulate *STM* mRNA and cells predicted to give rise to the SAM based on

histological analysis¹⁰.

It has been proposed that embryogenesis in *Arabidopsis* occurs in two parts¹⁶. In the first part, three regions are set up in the embryo: apical, central and basal. In the second part, this pattern is refined through interactions between the three basic compartments. One such refinement would be the subdivision of the apical region into cotyledon and SAM domains. In the framework of this model, *STM* is required for the development of the SAM domain only, its transcription occurring after the apical subdivision has occurred. It is as yet unknown what inter-compartment interactions would mediate this refinement.

The expression of *STM* mRNA persists into the seedling and adult plant where it is present in all four types of SAMs: vegetative, axillary, inflorescence and floral. In the vegetative meristem, *STM* mRNA accumulates in the SAM but is not found in leaves or leaf primordia (Fig. 3i). Domains of cells within the meristem also fail to accumulate *STM* mRNA (Fig. 3i). Transverse sections show that these domains are in positions at which leaves will arise (data not shown), suggesting that downregulation of *STM* is an early event in leaf formation. The persistence of *STM* expression in the meristem suggests that it is required for SAM maintenance as well as initiation. Although the most intense staining is in the SAM, *STM* mRNA accumulates in stem tissue and is therefore not uniquely restricted to the SAM (Figs 3i and 4a). The stem does not stain uniformly as cross-sections through the stem show a domain lacking *STM* expression around the vascular strands (data not shown). *STM* expression is not seen in *STM* mutant seedlings (Fig. 3j).

In the inflorescence meristem, *STM* expression disappears as floral buds are initiated (Fig. 4a,b). In this regard, the incipient floral buds behave like leaf primordia. However, *STM* mRNA

subsequently reappears in the floral meristem (Fig. 4b). Once re-established, *STM* expression is observed in the floral meristem as it generates floral organs. Early in the development of floral organs, as in leaves, *STM* expression is absent (Fig. 4c). However, as the carpels form, *STM* remains expressed medially, in two

ridges of tissue (Fig. 4d,e). Ovules will develop from the flanks of these ridges, indicating that they may act as 'meristems' for generating ovules.

The embryonic and vegetative pattern of *STM* expression is very similar to that of *KNOTTED1* in maize. (One difference is that

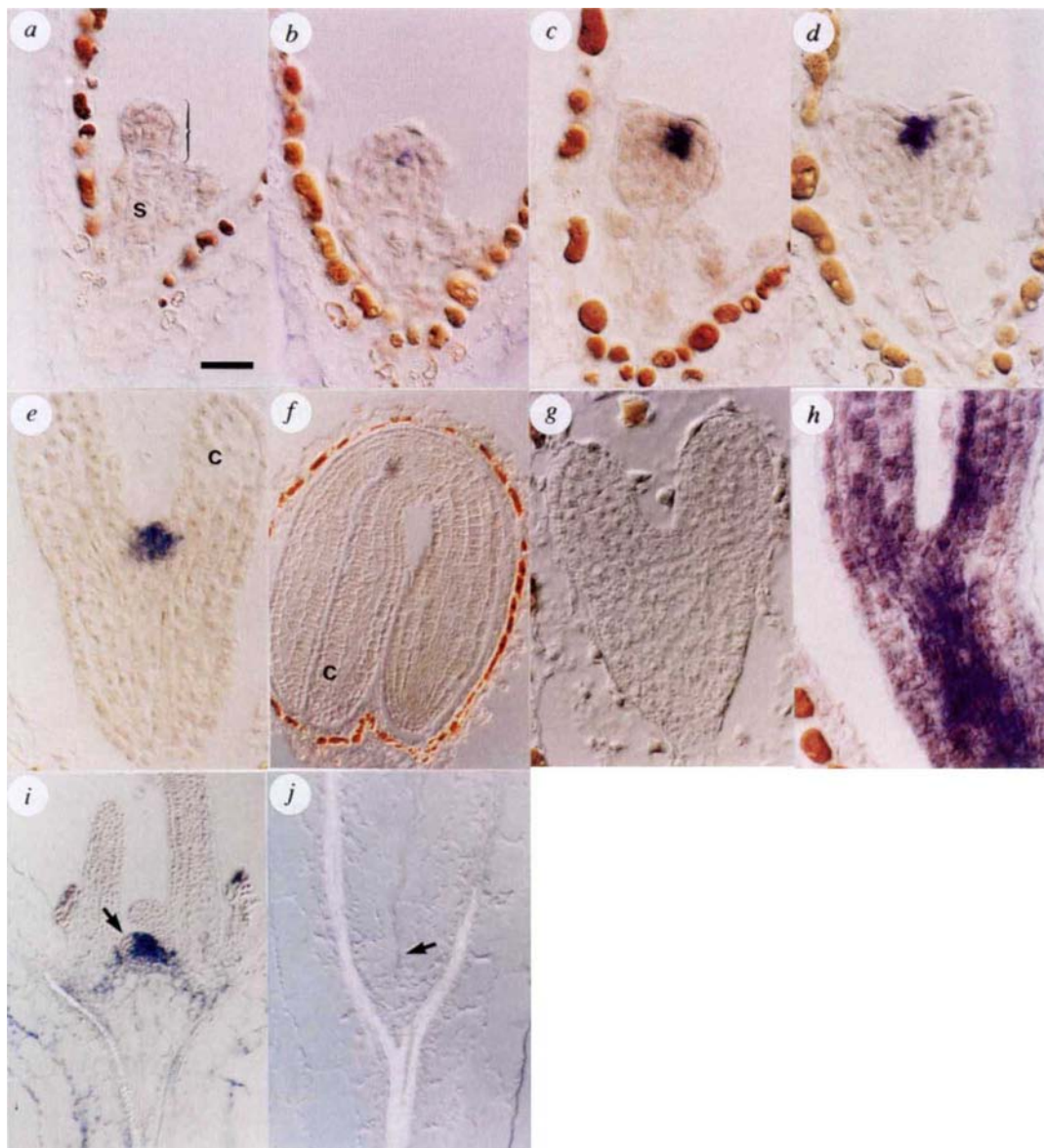


FIG. 3 Expression pattern of *STM* mRNA during wild-type embryogenesis and in the young seedling. *a*, Dermatogen-stage embryo. No *STM* expression is detected. Note that the integument takes on a brown colour in these figures. The origin of this colour is unknown but is not due to *STM* mRNA accumulation because it is present before hybridization. Bracket, embryo proper; gives rise to all of embryo except a portion of the root; *s*, suspensor. *b*, Early to mid-globular-stage embryo (32–64-cell stage). *STM* expression is first detected in a single cell in the center of the top half of the globular embryo. Some embryos at this stage show no detectable expression indicating that this represents the earliest onset of *STM* expression. *c*, Frontal view of late-globular-stage embryo. *STM* expression appears as a stripe running medially across the top of the embryo. Serial sections in the longitudinal plane perpendicular to this one confirm this interpretation. *d*, Heart-stage embryo. *STM* expression is confined to those cells in between the growing cotyledons that will give rise to the SAM. *e*, Torpedo-stage embryo. *f*, Mature-stage embryo. Note that *STM* expression is confined to the SAM, no expression is detected in the root apical meristem. *g*, Early torpedo-stage embryo probed with *STM* sense probe. *h*, Early bending cotyledon-stage embryo probed with antisense ubiquitin probe. *i*, Circa 1-

week-old wild-type seedling showing *STM* expression in the SAM. Note absence of expression in leaves and at site of incipient leaf primordium (arrow). *j*, Circa 1-week-old *stm-1* seedling. This seedling entirely lacks a SAM and also fails to show any expression of *STM* mRNA. Arrow indicates position normally occupied by the SAM. Scale bar, 25 μ m in *a*–*e*, *g*, *h*, and 77 μ m in *f*, *i* and *j*.

METHODS. *In situ* hybridization to mRNA was performed as in ref. 5. Stages of embryogenesis are as defined in ref. 18. Two antisense probes to *STM* were used. One spans the region from amino acids 81–186; the other spans from amino acids 81–382 and includes the 3'UTR. Both probes detect the same pattern of expression. Furthermore, cross-hybridization to either *KNAT1* or *KNAT2* mRNA is unlikely in these experiments because 1/*KNAT2* mRNA was not detectable under these conditions when the *KNAT2* gene itself was used as a probe in *in situ* hybridization experiments (C. Lincoln and S. Hake, unpublished observations) and 2/*KNAT1* shows a distinct pattern of hybridization from *STM*⁵. The sense probe spans the region from amino acids 81–382 and includes the 3'UTR. The ubiquitin antisense probe was synthesized from a 1.2-kb *Kpn* fragment of the *Arabidopsis* *UBQ4* gene¹⁹.

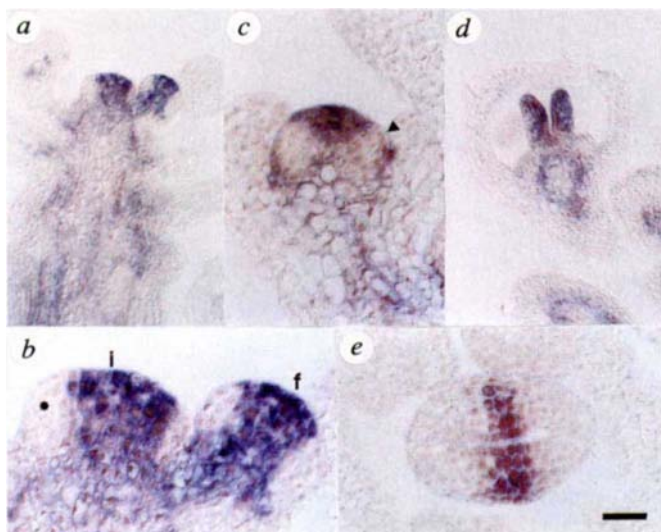


FIG. 4 Expression of *STM* during inflorescence and floral development. *a*, Inflorescence showing *STM* expression in central portions of inflorescence meristem and in central portion of stage 2 floral buds. *b*, Higher magnification of *a*. *STM* expression is downregulated in stage 1 floral buds (asterisk). Note lack of *STM* expression at position of incipient sepal primordia. *i*, Inflorescence meristem; *f*, stage 2 floral meristem. *c*, A stage 3 flower. *STM* expression is lacking at position of incipient medial stamen primordia (arrow). *d*, Stage 7–8 floral bud showing expression of *STM* in central portion of developing gynoecium. *e*, Transverse section through stage 7–8 gynoecium. Two ridges of expression are observed. Ovules will develop off the flanks of these ridges. Scale bar, 77 μ m in *a* and *d* and 25 μ m in *b*–*e*. Stages of flower development are as defined in ref. 20.

whereas *KN1* mRNA is not detectable in the maize L1 layer, *STM* mRNA is detectable in the *Arabidopsis* L1 layer.) It has been hypothesized that *KN1* functions to keep cells in an undifferentiated, undetermined, state^{6,7}. This is similar to the proposed role for *STM* in SAM maintenance and suggests that these genes may be orthologous to one another. □

Received 3 August; accepted 20 October 1995.

1. Vollbrecht, E., Veit, B., Sinha, N. & Hake, S. *Nature* **350**, 241–243 (1991).
2. Kerstetter, R. et al. *Pl. Cell* **6**, 1877–1887 (1994).
3. Ma, H., McMullen, M. D. & Finer, J. J. *Pl. molec. Biol.* **24**, 465–473 (1994).
4. Muller, K. J., Romano, N. & Rohde, W. *Nature* **374**, 727 (1995).
5. Lincoln, C., Long, J., Yamaguchi, J., Serikawa, K. & Hake, S. *Pl. Cell* **6**, 1859–1876 (1994).
6. Smith, L. G., Greene, B., Veit, B. & Hake, S. *Development* **116**, 21–30 (1992).
7. Jackson, D., Veit, B. & Hake, S. *Development* **120**, 405–413 (1994).
8. Freeling, M. & Hake, S. *Genetics* **111**, 617–634 (1985).

9. Becraft, P. W. & Freeling, M. *Genetics* **136**, 295–311 (1994).
10. Barton, M. K. & Poethig, R. S. *Development* **119**, 823–831 (1993).
11. Sinha, N. R., Williams, R. E. & Hake, S. *Genes Dev.* **7**, 878–795 (1993).
12. Mitchell, P. J., Urlaub, G. & Chasin, L. *Molec. cell. Biol.* **6**, 1926–1935 (1986).
13. Kuiper, M. T. R., Holthrop, M., Vennema, H., Lambowitz, A. M. & deVries, H. *J. biol. Chem.* **263**, 2848–2852 (1988).
14. Aroian, R. V. et al. *Molec. cell. Biol.* **13**, 626–637 (1993).
15. Mansfield, S. G. & Briarty, L. G. *Can. J. Bot.* **69**, 447–476 (1991).
16. Jürgens, G. *Cell* **81**, 467–470 (1995).
17. McConnell, J. & Barton, M. K. *Dev. Gen.* **16**, 358–366 (1995).
18. Jürgens, G. & Mayer, U. in *EMBRYOS Color Atlas of Development* (ed. Bard, J.) 7–22 (Wolfe, London, 1994).
19. Burke, T., Callis, J. & Vierstra, R. D. *Molec. gen. Genet.* **213**, 435–443 (1988).
20. *Arabidopsis: An Atlas of Morphology and Development* (ed. Bowman, J.) 133–172 (Springer, New York, 1994).

ACKNOWLEDGEMENTS. We thank P. Masson, C. Lincoln, S. Hake, J. Petrini and members of the Barton laboratory for helpful discussions and M. Evans and P. Masson for critical reading of the manuscript. R. Vierstra's laboratory kindly provided ubiquitin encoding plasmid p112. The *STM* sequence has been given GenBank accession number U32344 (Laboratory of Genetics paper no. 3441).

A role for glucagon-like peptide-1 in the central regulation of feeding

M. D. Turton, D. O'Shea, I. Gunn, S. A. Beak, C. M. B. Edwards, K. Meeran, S. J. Choi, G. M. Taylor, M. M. Heath, P. D. Lambert, J. P. H. Wilding, D. M. Smith, M. A. Ghatei, J. Herbert* & S. R. Bloom†

Endocrine Unit, Department of Medicine, Royal Postgraduate Medical School, Hammersmith Hospital, Du Cane Road, London W12 0NN, UK
* Department of Anatomy, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK

THE sequence of glucagon-like peptide-1 (7–36) amide (GLP-1) is completely conserved in all mammalian species studied, implying that it plays a critical physiological role¹. We have shown that GLP-1 and its specific receptors are present in the hypothalamus^{2,3}. No physiological role for central GLP-1 has been established. We report here that intracerebroventricular (ICV) GLP-1 powerfully inhibits feeding in fasted rats. ICV injection of the specific GLP-1-receptor antagonist, exendin (9–39)⁴, blocked the inhibitory effect of GLP-1 on food intake. Exendin (9–39) alone had no influence on fast-induced feeding but more than

doubled food intake in satiated rats, and augmented the feeding response to the appetite stimulant, neuropeptide Y. Induction of *c-fos* is a marker of neuronal activation⁵. Following ICV GLP-1 injection, *c-fos* appeared exclusively in the paraventricular nucleus of the hypothalamus and central nucleus of the amygdala, and this was inhibited by prior administration of exendin (9–39). Both of these regions of the brain are of primary importance in the regulation of feeding⁶. These findings suggest that central GLP-1 is a new physiological mediator of satiety.

We report that ICV administration of GLP-1 reduces food intake in fasted rats, with greater effect at higher doses (Fig. 1*b*). ICV injection of GLP-1 in rats at the beginning of the dark (feeding) phase also results in a profound decrease in feeding (Fig. 1*a*). When administered intraperitoneally up to a dose of 500 μ g, GLP-1 did not affect early dark-phase feeding (data not shown), suggesting that the action of GLP-1 on food intake is through central rather than peripheral mechanisms. A reduction in locomotor activity is a well defined part of the satiety sequence and follows nutrient ingestion⁷. In a subgroup of the animals given ICV GLP-1 at the beginning of the dark phase, locomotor activity was monitored by the frequency of line-crossing⁸. A significant reduction in activity was seen after ICV administration of GLP-1 (10 μ g; $41 \pm 7\%$ of control activity, $P < 0.05$; 100 μ g; $32 \pm 9\%$, $P < 0.01$, $n = 8$ per group) compared to controls. Following ingestion of a palatable meal⁹, the reduction in activity was similar to that observed following ICV injection of GLP-1 (10 μ g) (palatable meal; $54 \pm 19\%$ of control activity, $P < 0.05$, $n = 6$). Although not assessed formally, the behaviour of the GLP-1-treated animals could not be distinguished, by observation, from those fed a palatable meal¹⁰. Fragments of GLP-1 are inactive peripherally¹¹. To establish the specificity of GLP-1 on feeding,

† To whom correspondence should be addressed.

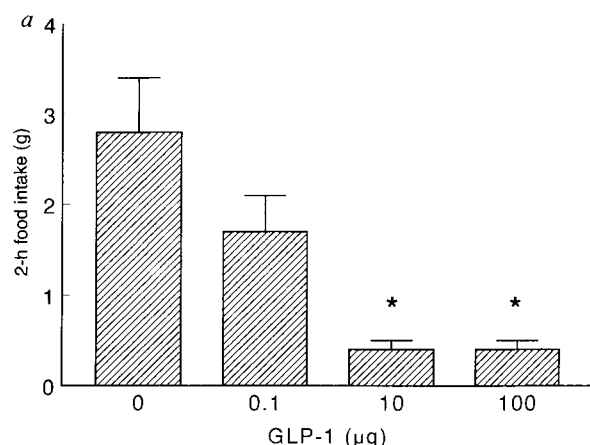
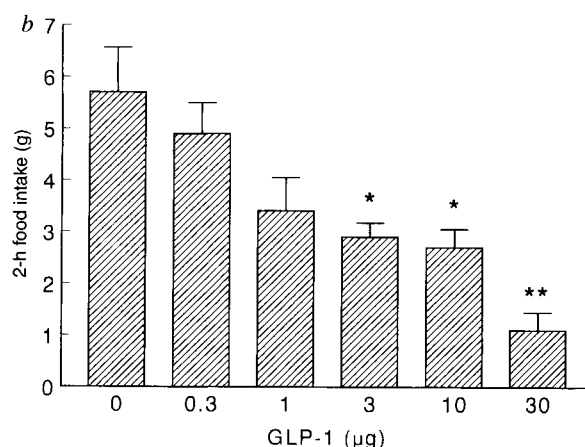


FIG. 1 a, Reduction in 2-h food intake in rats ($n = 7-8$ per group) injected ICV with GLP-1 or 0.9% saline ($10 \mu\text{l}$) at the onset of the dark phase ($F(3,27) = 10.6$, $P < 0.001$), (GLP-1 versus control, $*P < 0.001$). b, reduction in 2-h food intake in 24-h fasted rats ($n = 6-7$ per group) injected ICV with GLP-1 or 0.9% saline ($10 \mu\text{l}$) ($F(5,34) = 8.0$, $P < 0.001$), (GLP-1 versus control, $*P < 0.05$, $**P < 0.01$). METHODS. Adult male rats (200–250 g) were maintained in individual cages under controlled conditions of temperature ($21-23^\circ\text{C}$) and light (13-h light, 11-h dark) with *ad libitum* access to chow and water. Rats were anaesthetized and stainless steel guide cannulae implanted into the third cerebral ventricle as previously described¹⁷. After a 5-day recovery period, rats ($<10\%$) exhibiting a negative drinking response to ICV angiotensin II (All) ($150 \text{ ng } 3 \mu\text{l}^{-1}$) were excluded. Responders were sham injected before



the study, and weighed and handled daily. Substances were dissolved in 0.9% saline and administered ICV as previously described¹⁷. Following injection, rats were returned to cages containing pre-weighed chow and observed. At the end of the study period, chow was reweighed and total food intake in grams (g) for each rat calculated. All injections were given at least 72 h apart, between 08:30 and 11:00. Cannula placement was verified at the end of the study by a drinking response to ICV All, followed by ICV injection of dye, removal of the brain and visual examination of coronal slices. Those exhibiting a negative drinking response or absence of dye in the ventricular system were excluded. Results are shown as mean $\pm$ s.e.m. Comparisons between groups of data were made using *t*-tests or ANOVA with post-hoc Tukey's test.

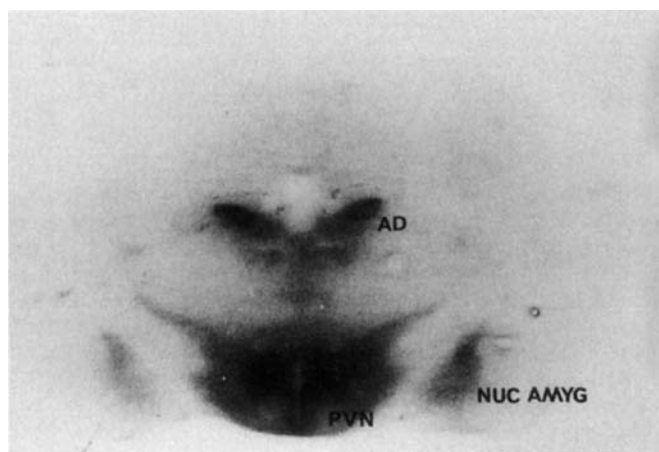


FIG. 2 Receptor autoradiography of GLP-1 binding sites in rat brain. Autoradiograph of a coronal section of rat brain incubated with 17 pM mono-¹²⁵I-GLP-1 alone, indicating total binding. No binding was observed in the presence of 200 nM unlabelled GLP-1 (nonspecific binding, not shown). High-density binding sites are visible in the hypothalamus (in particular the PVN), central nucleus of the amygdala (NUC AMYG) and anteroposterior thalamic nucleus (AD). No specific binding of GLP-1 was observed in the presence of 5 nM exendin (9-39) (not shown). METHODS. ¹²⁵I-GLP-1 labelled by the chloramine-T method³ was purified by reverse-phase HPLC (Waters C₁₈ Novapak, Millipore). The specific activity of the label, determined by radioimmunoassay, was 60 Bq fmol⁻¹. Receptor autoradiography was carried out in a similar manner to that previously described²². Coronal brain sections (15 µm) were pre-incubated in 25 mM HEPES assay buffer, pH 7.4, containing 2 mM MgCl₂, 0.1% bacitracin, 0.05% Tween 20 and 1% BSA, for 20 min. Excess liquid was drained from the slides, which were then incubated in assay buffer containing ¹²⁵I-labelled GLP-1 (1,000 Bq ml⁻¹) for 90 min at room temperature. Non-specific binding was determined in the presence of 200 nM unlabelled peptide. The slides were washed four times for 30 s in ice-cold assay buffer, rinsed in ice-cold distilled water and dried overnight at 4 °C under vacuum. Autoradiographs of the slides were developed after 10 days exposure at -80°C .

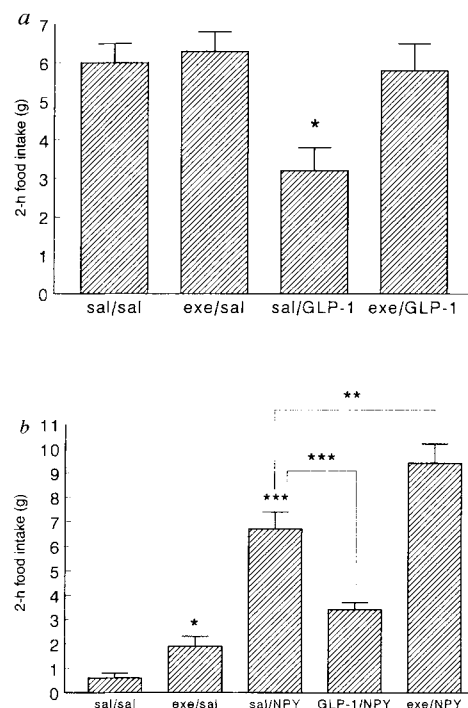


FIG. 3 Two-hour food intake in a, 24-h fasted rats ($n = 12-14$ per group) injected ICV with 0.9% saline (sal) ($5 \mu\text{l}$) or exendin (9-39) (exe) ($100 \mu\text{g } 5 \mu\text{l}^{-1}$), immediately followed by 0.9% saline ($5 \mu\text{l}$) or GLP-1 ($3 \mu\text{g } 5 \mu\text{l}^{-1}$) ($F(3,48) = 6.2$, $P < 0.005$; GLP-1 versus control, $*P < 0.05$). b, Rats fed *ad libitum* ($n = 6-9$ per group) injected ICV at the start of the light phase with 0.9% saline ($5 \mu\text{l}$) (sal) or NPY ($10 \mu\text{g } 5 \mu\text{l}^{-1}$), immediately followed by either 0.9% saline ($5 \mu\text{l}$), GLP-1 ($10 \mu\text{g } 5 \mu\text{l}^{-1}$) or exendin (9-39) (exe) ($100 \mu\text{g } 5 \mu\text{l}^{-1}$). Both exendin (9-39) and NPY significantly increased food intake ($*P < 0.05$ and $***P < 0.001$, respectively, compared to control). GLP-1 significantly reduced ($***P < 0.001$) whereas exendin (9-39) significantly increased ($**P < 0.01$) the feeding response to NPY. Methods as for Fig. 1.

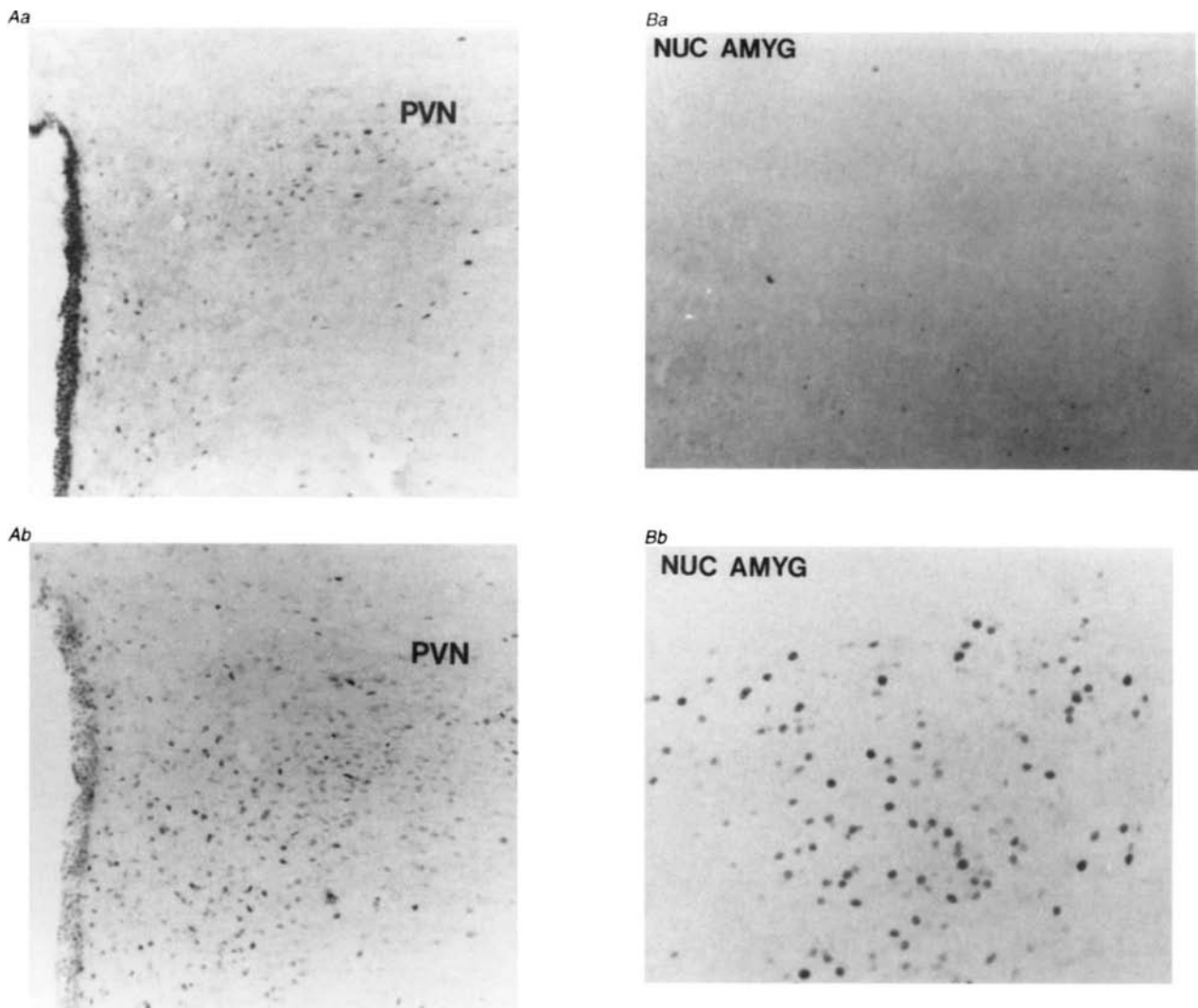


FIG. 4 Activation of *c-fos* following ICV injection of: Aa, 0.9% saline (10 μ l) (control), demonstrating standard basal expression in the PVN (21 ± 2.2 positive cells per $10^4 \mu\text{m}^2$, identical to that observed following ICV exendin (9-39) ($100 \mu\text{g } 10 \mu\text{l}^{-1}$) (not shown); and b, GLP-1 ($10 \mu\text{g } 10 \mu\text{l}^{-1}$). Significant activation is observed in the PVN (57 ± 2.3 positive cells per $10^4 \mu\text{m}^2$, $P < 0.001$ compared to control). Ba, 0.9% saline (10 μ l) demonstrating standard basal expression in the central nucleus of the amygdala (4 ± 0.5 positive cells per $10^4 \mu\text{m}^2$); and b, GLP-1 ($10 \mu\text{g } 10 \mu\text{l}^{-1}$). Significant activation is observed in the central nucleus of the amygdala (NUC AMYG) (13 ± 0.5 positive cells per $10^4 \mu\text{m}^2$; $P < 0.001$ compared to control).

METHODS. Adult male rats (200–250 g) were maintained in individual cages under controlled conditions of temperature ($21\text{--}23^\circ\text{C}$) and light (13-h light, 11-h dark) with *ad libitum* access to chow and water. Rats were anaesthetized and stainless steel guide cannulae were implanted into the

lateral cerebral ventricle as previously described²³. Rats were handled for 10 min each day following surgery, to familiarize them with infusion procedures and to minimize stress, which can readily activate *c-fos*²⁴. Substances were dissolved in 0.9% saline and administered ICV as previously described²³. Following injection, rats were returned to their cages for 1 h with no access to food. Rats were anaesthetized and perfused transcardially with 0.1 M PBS followed by 4% paraformaldehyde solution. The brains were removed and stored in 20% sucrose solution before being processed for detection of *c-fos* as previously described²³. Coronal sections (40 μm) were washed and incubated with antisera and processed using the Vectastain ABC kit (Vector Laboratories, USA). Sections were immersed in 0.1% diaminobenzidine tetrahydrochloride solution and mounted on slides. Expression of *c-fos* was then quantified as previously described²³. Results are expressed as mean $\pm$ s.e.m. of positive cells per $10^4 \mu\text{m}^2$. Comparisons between groups of data were made using *t*-tests.

GLP-1 (8–36), (9–36) and (11–36) amide were therefore tested by ICV injection in fasted rats. At doses of up to 100 μg , these peptides had no effect on food intake or locomotor activity compared to controls (data not shown).

Exendin (9-39) has sequence homology to GLP-1 (ref. 12). It is a highly selective antagonist at GLP-1 receptors *in vitro*⁴. Furthermore, we have shown that exendin (9-39) specifically abolishes GLP-1-induced insulin secretion in the rat at a ratio of 15:1 (ref. 13), and *in vitro* in a clonal β -cell line at a ratio of 10:1 (ref. 14), indicating blockade of β -cell GLP-1 receptors. We therefore used a dose of 100 μg of exendin (9-39) to antagonize a dose of 3 μg GLP-1 *in vivo*. Autoradiography of GLP-1 binding sites in the brain shows dense specific binding of GLP-1 in the hypothalamus, especially the paraventricular nucleus (PVN); the central nucleus of the amygdala

and the anterodorsal thalamic nucleus (Fig. 2). Addition of 5 nM exendin (9-39) blocks GLP-1 binding in these regions (not shown).

ICV administration of exendin (9-39) (100 μg) blocks the inhibitory effect of GLP-1 (3 μg) on food intake in fasted rats (Fig. 3a), but exendin (9-39) alone does not affect food intake in this circumstance (Fig. 3a). When administered by ICV injection to rats fed *ad libitum* at the start of the light phase (satiated), exendin (9-39) (100 μg) more than doubled food intake (Fig. 3b). The profound effect of exendin (9-39) in satiated animals without effect in fasted animals suggests a role for endogenous GLP-1 as a central satiety factor¹⁵.

Neuropeptide Y (NPY) is the most powerful stimulant of feeding known^{16,17}. In animals fed *ad libitum*, at the start of the light phase, ICV administration of GLP-1 (10 μg) immediately

before NPY (10 µg) greatly reduced food intake (Fig. 3b). Furthermore, ICV administration of exendin (9-39) (100 µg) immediately before NPY (10 µg) significantly increased food intake, compared to treatment with NPY (10 µg) alone (Fig. 3b). Hypothalamic NPY messenger RNA was assessed, as previously described¹⁸, following repeated ICV administration of 0.9% saline or GLP-1 (100 µg) in both 72-h fasted and *ad libitum* fed rats. ICV injections were given during this 72-h period at 48, 24, 12 and 4 h before being killed. No change in NPY mRNA was found following ICV GLP-1 injection (data not shown). This suggests that GLP-1 does not act by altering hypothalamic NPY synthesis. The increase in food intake following blockade of GLP-1 receptors by exendin (9-39) and the augmented NPY response with coadministration of exendin (9-39) supports a physiological role for central GLP-1 in the regulation of feeding.

To establish the neuronal population(s) activated by central GLP-1, expression of the immediate early gene *c-fos*, a well established marker of neuronal activation⁵, was mapped following ICV injection of GLP-1. Administration of exendin (9-39) (data not shown) or 0.9% saline (Fig. 4Aa, Ba) did not activate *c-fos* in any region of the brain. In contrast, rats injected with GLP-1 exhibit dense expression of *c-Fos* in the PVN (Fig. 4Ab) and central nucleus of the amygdala (Fig. 4Bb), but not the remainder of the hypothalamus or anterodorsal thalamic nucleus. The induction of *c-fos* following ICV GLP-1 injection was inhibited by prior administration of exendin (9-39) (data not shown).

In conclusion, central GLP-1 is a new physiological regulator of feeding in the rat. Many neuropeptides have been proposed as central satiety factors⁶, but few have an established physiological role. GLP-1 may be the most potent inhibitor of feeding yet identified in the rat. Its role in man remains to be determined.

The interaction of GLP-1 with other established regulators of food intake, such as leptin (Ob protein)^{19,21} and NPY, also needs investigation. Understanding the function of these centrally acting regulators should ultimately lead to new and more effective agents for the management of appetite disorders. □

Received 2 October; accepted 3 November 1995.

- Orskov, C. *Diabetologia* **35**, 701–711 (1992).
- Kreymann, B. et al. *Brain Res.* **502**, 325–331 (1989).
- Kanse, S. M., Kreymann, B., Ghatel, M. A. & Bloom, S. R. *FEBS Lett.* **241**, 209–212 (1988).
- Goke, R. et al. *J. Biol. Chem.* **268**, 19650–19655 (1993).
- Sagar, S. M., Sharp, F. R. & Curran, T. *Science* **240**, 1328–1331 (1988).
- Morley, J. E. *Endocr. Rev.* **8**, 256–287 (1987).
- Antin, J., Gibbs, J., Holt, J., Young, R. C. & Smith, G. P. *J. comp. Physiol. Psychol.* **89**, 784–790 (1975).
- Chow, H. L. & Beck, C. H. *Eur. J. Pharmac.* **102**, 297–304 (1984).
- Wilding, J. P. et al. *J. Endocr.* **132**, 299–304 (1992).
- Blundell, J. E., Rogers, P. J. & Hill, A. J. *Brain Res. Bull.* **15**, 371–376 (1985).
- Suzuki, S., Kawai, K., Ohashi, S., Mukai, H. & Yamashita, K. *Endocrinology* **125**, 3109–3114 (1989).
- Rai, A., Singh, G., Raffanelli, R., Eng, J. & Raufman, J. P. *Am. J. Physiol.* **265**, G118–125 (1993).
- Wang, Z. et al. *J. clin. Invest.* **95**, 417–421 (1995).
- Kulkarni, R. N. et al. *Reg. Pept.* **57**, 201 (1995).
- Billington, C. J., Levine, A. S. & Morley, J. E. *Am. J. Physiol.* **245**, R920–926 (1983).
- Clark, J. T., Kalra, P. S. & Kalra, S. P. *Endocrinology* **117**, 2435–2442 (1985).
- Lambert, P. D. et al. *Endocrinology* **133**, 29–32 (1993).
- Wilding, J. P., Gilbey, S. G., Lambert, P. D., Ghatel, M. A. & Bloom, S. R. *Neuroendocrinology* **57**, 581–587 (1993).
- Pelleymounter, M. A. et al. *Science* **269**, 540–543 (1995).
- Halaas, J. L. et al. *Science* **269**, 543–546 (1995).
- Campfield, L. A., Smith, F. J., Guise, Y., Devos, R. & Burn, P. *Science* **269**, 546–549 (1995).
- Veale, P. R., Bhogal, R., Morgan, D. G., Smith, D. M. & Bloom, S. R. *Eur. J. Pharmac.* **262**, 133–141 (1994).
- Lambert, P. D., Phillips, P. J., Wilding, J. P. H., Bloom, S. R. & Herbert, J. *Brain Res.* **670**, 59–65 (1995).
- Arnold, F. J. et al. *Neuroscience* **51**, 377–390 (1992).

ACKNOWLEDGEMENTS. We thank P. Byfield for the synthesis of all peptides used in these studies and P. Phillips for his expert technical assistance.

Rapid colour-specific detection of motion in human vision

Simon J. Cropper* & Andrew M. Derrington†

* School of Psychology, University of Wales College of Cardiff, Cardiff CF1 3YG, UK

† Department of Psychology, University Park, Nottingham NG7 2RD, UK

THE human visual system is much better at analysing the motion of luminance (black and white) patterns than it is at analysing the motion of colour patterns^{1–4}, especially if the pattern is presented very briefly⁵ or moves rapidly⁶. We report here that observers reliably distinguish the direction of motion of a colour pattern presented for only 17 milliseconds, provided that the contrast is several times the threshold value (the contrast needed to detect the presence of the pattern). A control experiment, in which a static luminance 'mask' is added to the moving colour pattern, proves that discrimination of the direction of motion of these brief stimuli is colour-specific. The mask drastically impairs discrimination of the direction of motion of a luminance pattern, but it has little effect on a colour pattern. We conclude that the human visual system contains colour-specific motion-detection mechanisms that are capable of analysing very brief signals.

The prevailing view is that the visual system's primary analyses of colour and movement occur in parallel in different areas of the brain^{7–9}. An important observation supporting this view is that the motion of many patterns defined solely by colour is undetectable¹ or hard to see⁴. The colour of such a pattern must be analysed before its motion, so if motion and colour are analysed in parallel, motion analysis must inevitably be compromised.

To demonstrate that colour contributes directly to the analysis

of motion, it is important to show that motion discrimination is not secondary to analysis of position¹⁰ or to the production of eye movements that track the stimulus. The only way to be sure of this is to use a very brief presentation so that the stimulus disappears before an eye movement can be initiated. Previous demonstrations that colour contributes distinctively to motion analysis have always used long presentations^{11,12}. Thus, even though direction-discrimination thresholds for colour stimuli may be lower (in cone contrast) than those of luminance stimuli⁶, we cannot yet exclude the possibility that the motion of such stimuli is analysed by conscious monitoring of changes in position¹⁰.

Figure 1 shows that the motion of colour gratings can be discriminated reliably even when the grating is only present for 17 ms, about 5 times less than the time taken to initiate an eye movement¹³. Discrimination of such brief stimuli could only be mediated by a mechanism whose primary function is the analysis of motion.

We must ensure that the motion of colour gratings shown in Fig. 1 depends on a mechanism selective for colour and is not caused by inadvertent stimulation of mechanisms sensitive to luminance patterns¹⁴. Accordingly, we retested motion discrimination using a red–green colour grating to which a static luminance mask had been added. The test depends on the assumption that if the colour grating's motion is detected by a luminance-sensitive mechanism, then we can treat it as equivalent to the luminance grating that would support the same direction-discrimination performance. The exact position of the luminance grating relative to the colour grating (its phase) is unknown because it depends on the nature of the hypothetical artefact that causes the colour grating to stimulate a luminance-sensitive mechanism.

The basis of the masking test is shown in Fig. 2, which illustrates the effect of static masks added to a moving luminance grating in different phases. In the absence of the mask (lower left-hand

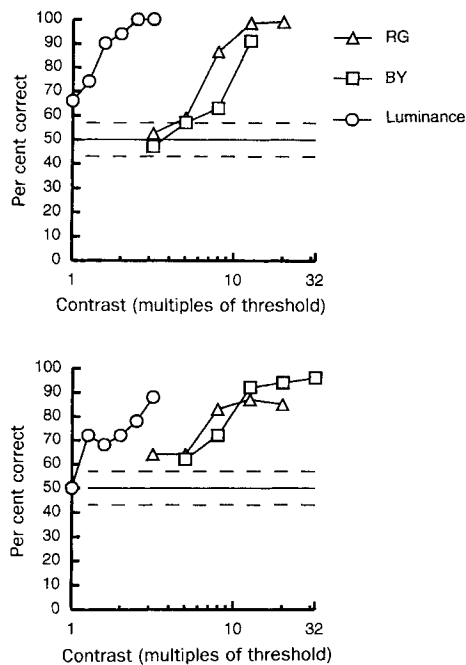
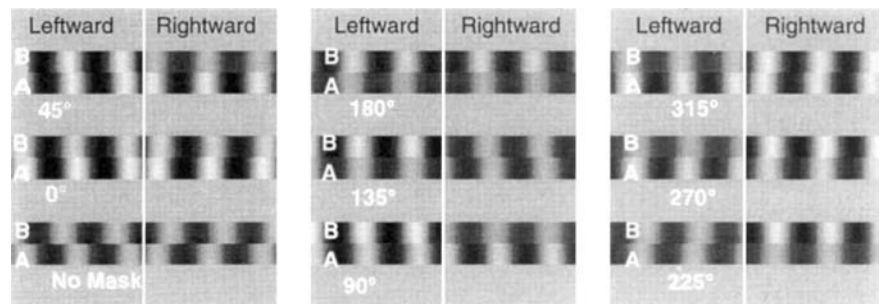


FIG. 1 Direction discrimination by two observers, S.J.C. (top) and N.M. (bottom), as a function of contrast (relative to detection threshold) in a temporal two-alternative forced choice task using luminance (circles) or colour (triangles, squares) gratings. Dashed lines indicate 95% confidence intervals for chance performance. Observer N.M. is naive. Horizontal 1 cycle per degree sinusoidal modulations of the luminance or colour of a 120-Hz television monitor (Mitsubishi HL7955) moved suddenly 0.25° upwards or downwards in the middle of a presentation lasting two frames (16.7 ms). On each trial the same pattern was presented twice, moving in opposite directions in an order selected at random, and the subject indicated whether it had moved upwards on the first or second presentation. The colour modulations are those that best stimulate the two classes of parvocellular neuron. Red-green (triangles) is the 0° cardinal direction that modulates L and M cones in antiphase¹⁹; it does not stimulate S cones. Blue-yellow (squares) is the 90° direction that only stimulates S cones^{16,19}.

FIG. 2 Each of the nine panels shows the two frames (A and B) of an apparent motion stimulus which jumps leftwards in the left-hand half of the panel and rightwards in the right-hand half. Frame A is the first frame and frame B is the second frame of the moving stimulus. The bottom left-hand panel shows a high-contrast unmasked stimulus. The remaining panels show a low-contrast stimulus masked by a static grating three times its contrast. Mask phase is given in degrees on the left-hand panel. With no mask, the motion can be seen clearly: frame B is to the left of frame A in the left-hand half and to the right of it in the right-hand half. The mask makes the motion more



difficult to see at all phases, and at some phases (90° , 135° , 180°) it reverses the direction of one or both of the moving stimuli.

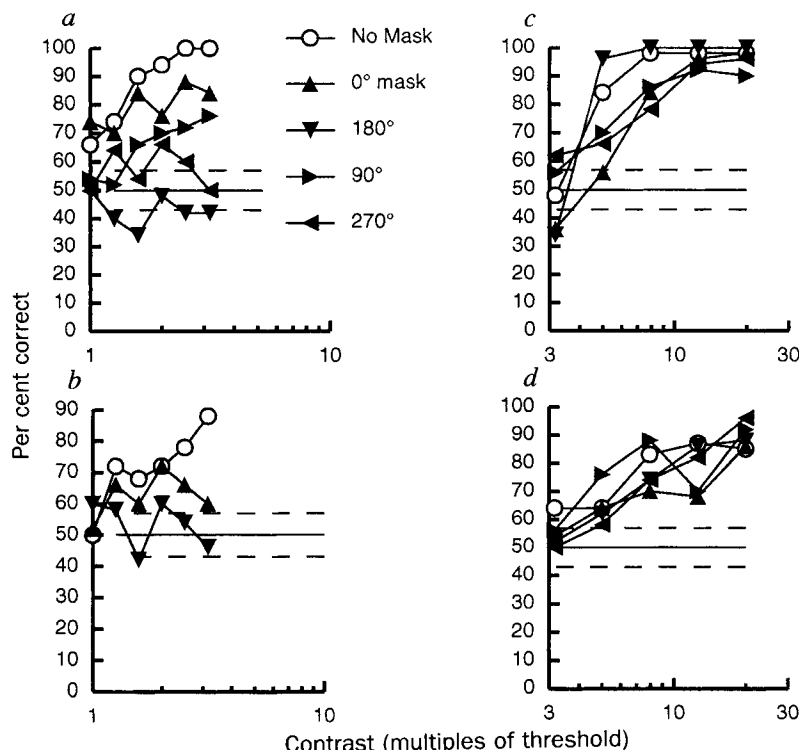


FIG. 3 a, b, Direction discrimination by S.J.C. and N.M., respectively, for a luminance grating presented alone or added to a static luminance mask of different phases. c, d, Similar results obtained using a moving colour grating and a static luminance mask (c, S.J.C.; d, N.M.). The mask contrast was $\sqrt{10}$ -times detection threshold. The mask coincided exactly in time with the stimulus. Other details are as described for Fig. 1.

panel), the grating in the left-hand half of the panel moves leftwards by half the width of a bar between frame A and frame B; the grating in the right-hand half of the panel moves rightward by the same distance.

Adding the stationary mask to the moving grating produces a pattern in which the motion is reduced, and may be reversed, depending on the phase of the mask. When the mask and frame A of the moving grating are in phase, the size of the step is smaller, and the motion is accompanied by a change in contrast. At several other phases (for example, 90° , 135° and 180°), the mask reverses one or both of the directions of motion. These effects are clearest when the mask has higher contrast than the test. So we can expect that if the colour grating's motion is signalled by a luminance-sensitive mechanism, static luminance masks will substantially impair direction-discrimination performance near threshold, and that the severity of the impairment will depend on the phase of the mask.

Figure 3 shows, for two observers S.J.C. and N.M., that a luminance mask disrupts performance in discriminating the direction of motion of a luminance grating, but has little effect when the target is a colour grating. When the mask is added in phase with the luminance grating (phase 0°), it reduces performance slightly in one observer and substantially in the other. At 180° phase, the mask reduces performance severely over a wide range of contrasts.

When the moving stimulus is a colour grating (Fig. 3c and d), the luminance mask causes a modest reduction in performance in some phases and has no effect in others. Evidently the moving colour pattern is sensed by a mechanism that is almost completely insensitive to luminance variations. The modest reductions in performance might reflect a residual luminance sensitivity in a red-green opponent mechanism caused by a phase difference between L and M cones¹⁵.

In conclusion, the human visual system must contain a colour-selective motion analyser that is capable of dealing with very brief stimuli. The neural substrate must almost certainly involve parvocellular neurons, because the sensitivity of magnocellular neurons to colour is so much worse than their sensitivity to luminance¹⁶ that it is unlikely that they would be selective for colour. The motion of our colour gratings would not be signalled by a mechanism sensitive to the motion of unsigned chromatic borders, like that putatively based on magnocellular neurons¹⁷, because the 0.25-cycle jump would give a completely ambiguous signal. The results we describe here may be a psychophysical correlate of the chromatic motion signals recorded in cortical area MT (ref. 18). □

Received 24 May; accepted 17 October 1995.

1. Ramachandran, V. S. & Gregory, R. L. *Nature* **275**, 55–56 (1978).
2. Cavanagh, P., Tyler, C. W. & Favreau, O. E. *J. opt. Soc. Am.* **A1**, 893–899 (1984).
3. Cavanagh, P. & Anstis, S. *Vision Res.* **31**, 2109–2148 (1991).
4. Mullen, K. T. & Boulton, J. C. *Vision Res.* **32**, 483–488 (1992).
5. Cropper, S. J. & Derrington, A. M. *Vision Res.* **34**, 49–58 (1994).
6. Derrington, A. M. & Henning, G. B. *Vision Res.* **33**, 799–811 (1993).
7. Zeki, S. M. *J. Physiol., Lond.* **277**, 273–290 (1978).
8. Zeki, S. M. *Nature* **274**, 423–428 (1978).
9. Livingstone, M. S. & Hubel, D. H. *J. Neurosci.* **7**, 3416–3468 (1987).
10. Cavanagh, P. *Science* **275**, 1563–1565 (1992).
11. Gorea, A., Papathomas, T. V. & Kovacs, I. *Vision Res.* **33**, 2515–2534 (1993).
12. Hawken, M. J., Gegenfurtner, K. R. & Tang, C. *Nature* **367**, 268–270 (1994).
13. Merrison, A. F. A. & Carpenter, R. H. S. *Vision Res.* **35**, 1459–1462 (1995).
14. Troscianko, T. & Fahle, M. *J. opt. Soc. Am.* **A5**, 871–880 (1988).
15. Stromeyer, C. F., Kronauer, R. E., Ryu, A., Chaparro, A. & Eskew, R. T. *J. Physiol., Lond.* **485**, 221–243 (1995).
16. Derrington, A. M., Krauskopf, J. & Lennie, P. *J. Physiol., Lond.* **357**, 241–265 (1984).
17. Dobkins, K. R. & Albright, T. D. *Vision Res.* **33**, 1019–1036 (1993).
18. Saito, H., Tanaka, K., Isono, H., Yasuda, M. & Mikami, A. *Exp Brain Res.* **75**, 1–14 (1989).
19. Krauskopf, J., Williams, D. R. & Heeley, D. W. *Vision Res.* **22**, 1123–1131 (1982).

ACKNOWLEDGEMENTS. We are grateful for financial support from the Science and Engineering Research Council, the Medical Research Council and the Wellcome Trust.

Activation of K^+ channels and suppression of neuronal activity by secreted β -amyloid-precursor protein

Katsutoshi Furukawa*, Steven W. Barger*, Eric M. Blalock† & Mark P. Mattson*‡§

* Sanders-Brown Research Center on Aging, † Department of Pharmacology, ‡ Department of Anatomy and Neurobiology, University of Kentucky, Lexington, Kentucky 40536, USA

THE Alzheimer's β -amyloid precursor protein (β -APP) is widely expressed in neural cells, and in neurons secreted forms of β -APP (sAPPs) are released from membrane-spanning holo- β -APP in an activity-dependent manner^{1,2}. Secreted APPs can modulate neurite outgrowth, synaptogenesis, synaptic plasticity and cell survival^{3–9}; a signal transduction mechanism of sAPPs may involve modulation of intracellular calcium levels ($[Ca^{2+}]_i$)^{4,10}. Here we use whole-cell perforated patch and single-channel patch-clamp analysis of hippocampal neurons to demonstrate that sAPPs suppress action potentials and hyperpolarize neurons by activating high-conductance, charybdotoxin-sensitive K^+ channels. Activation of K^+ channels by sAPPs was mimicked by a cyclic GMP analogue and sodium nitroprusside and blocked by an antagonist of cGMP-dependent kinase and a phosphatase

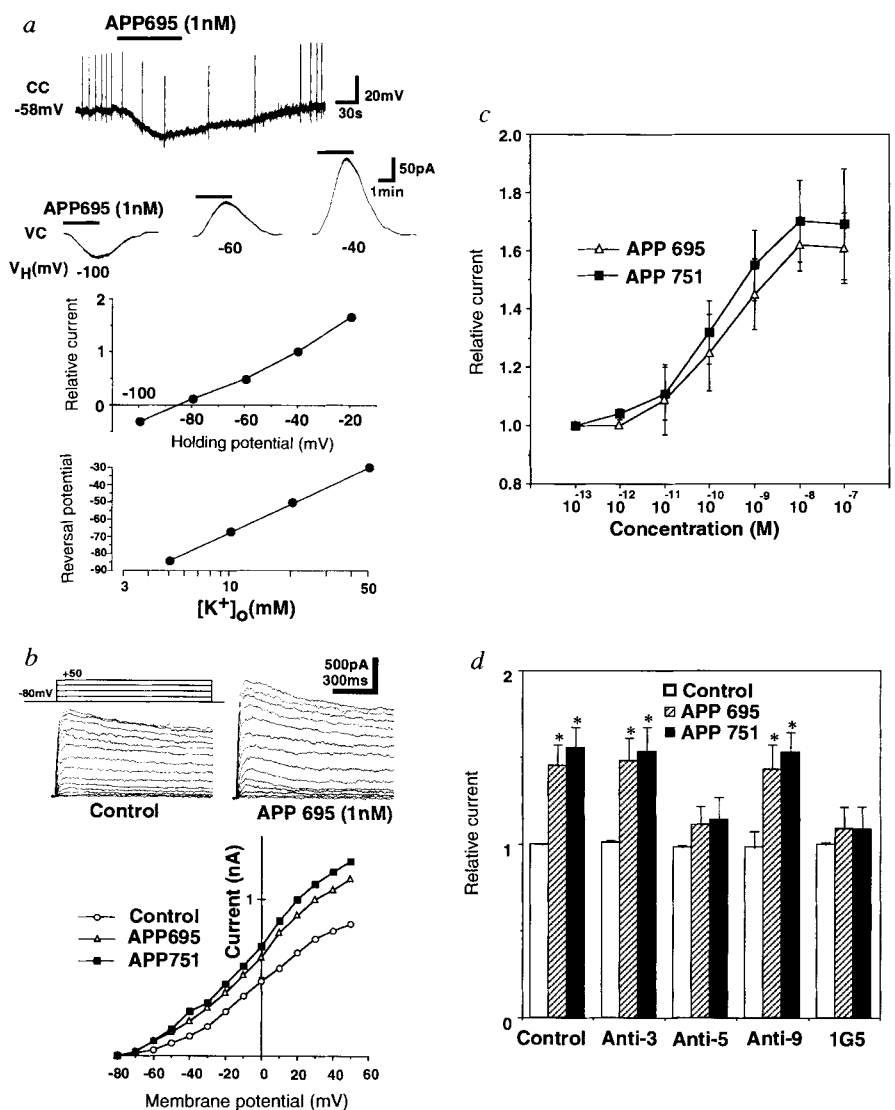
inhibitor, suggesting that the effect is mediated by cGMP and protein dephosphorylation. Calcium imaging studies indicate that activation of K^+ channels mediates the ability of sAPPs to decrease $[Ca^{2+}]_i$. Modulation of neuronal excitability may be a major mechanism by which β -APP regulates developmental and synaptic plasticity in the nervous system.

As described previously¹¹, cultured embryonic hippocampal neurons were synaptically connected and fired action potentials in response to above-threshold stimulation from synaptic inputs. Neurons had a resting membrane potential of -67.4 ± 4.5 mV ($n = 14$) under current-clamp using the nystatin perforated patch protocol; application of 1 nM sAPP695 induced a reversible hyperpolarization (Fig. 1a). Action potential frequency was reduced during application of 1 nM sAPP (frequency was 62.0 ± 18.8 spikes per 10 s before exposure to sAPP and 11.8 ± 3.3 spikes per 10 s following exposure to sAPP; $n = 5$; $P < 0.01$; paired t -test) and recovered to basal frequency 3–5 min after washout of sAPP (Fig. 1a). Under voltage-clamp conditions, sAPP695 evoked an outward current at holding potentials from -80 to -20 mV, with a reversal potential of -85 mV, which is essentially identical to the K^+ equilibrium potential (Fig. 1a). The reversal potential changed by 57 mV for a tenfold change in $[K^+]_o$ (Fig. 1a), and the sAPP-induced outward current was completely blocked by 10 mM tetraethylammonium (TEA; $n = 4$) and 30 nM charybdotoxin (CTX; Fig. 2), demonstrating that K^+ channels were responsible for the effect. The peak increase in outward current occurred within 60–90 s of exposure to sAPP, was maintained during continued sAPP exposure, and recovered to baseline following removal of sAPP. Voltage-activated I_K was increased to $\sim 150\%$ of control level at step pulses from -80 to $+50$ mV by either sAPP695 or sAPP751; this enhancement of I_K appeared to be relatively independent of

§ To whom correspondence should be addressed.

FIG. 1 Secreted APPs suppress spontaneous action potentials, hyperpolarize the membrane and activate potassium currents in cultured hippocampal neurons. **a**, Upper record shows spontaneous action potentials recorded in current clamp (CC) mode in a neuron before, during and following exposure to sAPP695. The sAPP elicited a transient hyperpolarization and greatly reduced firing rate. The other three traces show responses to 1 nM sAPP695 under voltage clamp (VC) at three different potentials; at -60 and -40 mV the sAPP elicited large outward currents. Similar results were obtained with sAPP695 ($n = 7$) and sAPP751 ($n = 5$). Top graph shows I/V relationship of sAPP-induced currents (reversal potential, -85 mV); lower graph shows the dependence of the reversal potential on the $[K^+]_o$ (mM). **b**, Top, representative current recordings in a neuron before (control) and 5 min after exposure to sAPP751. Pulses were applied in voltage steps from -80 to $+50$ mV. Bottom, I/V relationships of individual recordings in the same cell before treatment, 5 min after exposure to sAPP695, and 5 min after exposure to sAPP695 (8 min wash between exposures to sAPPs). **c**, Concentration-response relationships of sAPPs. The relative current amplitude (90 s after exposure to sAPP) was plotted against individual concentrations. Mean and s.e.m. ($n = 7-10$). **d**, The effects of domain-specific β -APP antibodies on K^+ current activated by sAPPs. Initial K^+ current was recorded, then neurons were exposed to $2.5 \mu\text{M ml}^{-1}$ of antibody alone (control) or a mixture of antibody ($2.5 \mu\text{M ml}^{-1}$) and sAPP (1 nM). Anti-5 and 1G5 recognize amino acids 444-592 of APP695; anti-3 and anti-9 recognize the N terminus and Kunitz protease-inhibitor domain of sAPPs, respectively. Values represent current relative to pretreatment level (mean and s.e.m., $n = 5-7$). Current was significantly increased by sAPP695 and sAPP751 in the presence of anti-3 and anti-9 ($P < 0.01$), but not in the presence of anti-5 or 1G5. **METHODS**. Dissociated cell cultures of fetal rat hippocampus (embryonic day 18) were established and maintained as described²⁸. Experiments were done in cells cultured for 7-14 d. Responses were recorded using a nystatin-perforated patch recording configuration²⁹. Under the culture conditions used, neurons were synaptically connected and fired action potentials in response to above-threshold stimulation from synaptic inputs; action potentials were recorded under current-clamp conditions as described¹². The resistance between the recording electrode filled with the internal solution and the reference electrode in the bath was 4-8 M Ω . The current was measured using a patch-clamp amplifier (Axopatch-1D).

membrane potential (Fig. 1b). The increase in outward current was not the result of a decrease in Na^+ or Ca^{2+} current because it was observed at depolarizing voltages greater than $+30$ mV. Bicuculline and atropine ($10 \mu\text{M}$ each) did not affect activation of I_K by sAPPs ($n = 3$), indicating that effects of sAPPs on GABA (γ -aminobutyric acid) or muscarinic receptors were not involved in the hyperpolarizing actions of sAPPs. Activation of I_K by sAPPs was concentration-dependent (10 pM to 10 nM), with similar curves for sAPP695 and sAPP751 (Fig. 1c). As with regulation of $[\text{Ca}^{2+}]_i$ by sAPPs^{3,10}, the potentiation of I_K by sAPPs was blocked by antibodies directed against a carboxy-terminal region of sAPP, but was unaffected by antibodies that recognize the amino terminus or Kunitz protease-inhibitor domain of sAPP (Fig. 1d).



The ionic composition of the standard external solution was (in mM) 150 NaCl, 5 KCl, 1 MgCl_2 , 2 CaCl_2 , 10 glucose, 10 HEPES, and $0.3 \mu\text{M}$ tetrodotoxin; pH was adjusted to 7.4 with Trizma base. The internal solution for the nystatin-perforated patch was 150 KCl and 10 HEPES; nystatin was dissolved in the internal solution at $50-100 \text{ mg ml}^{-1}$ just before use. The internal solution for the outside-out single channel recording was 15 KCl, 115 K-gluconate, 5 MgCl_2 , 5 EGTA, 10 HEPES and 5 $\text{Na}_2\text{-ATP}$, pH 7.2. Cells were constantly perfused with external solution and test agents were applied to the neuron via a Y-tube apparatus. Nonspecific leak currents induced by voltage steps were subtracted by applying small negative step pulses using the LeakSubtract program of pClamp (version 6.0.2; Axon Instruments). Secreted forms of APP (sAPP695 and sAPP751) were purified from the culture supernatant of HEK293 cells transfected to express either sAPP695 or sAPP751 as described¹¹.

Basal I_K was reduced by $\sim 30\%$ in neurons exposed to glibenclamide and was unaffected by CTX and apamin (Fig. 2a). The activation of I_K by sAPP was completely blocked by CTX but was unaffected by apamin. sAPPs activated I_K in the presence of glibenclamide, although the amplitude of the current was less than that seen in the absence of glibenclamide. In single-channel recordings, sAPP increased the open probability of a high conductance (~ 240 pS) K^+ channel and the current was suppressed by CTX (Fig. 2b). Glibenclamide did not affect the sAPP-induced increase in single-channel openings under our conditions (Fig. 2c), although an effect of sAPPs on the K_{ATP} channel cannot be ruled out. The conductance and pharmacological profile of the high-conductance, Ca^{2+} -activated K^+ channels activated by sAPPs

FIG. 2 Pharmacological characterization of sAPP-induced K^+ currents. *a*, Whole-cell current amplitude was recorded before and 90 s after exposure to the indicated agents; values represent the relative current post-treatment (mean and s.e.m.; $n = 5-9$). The concentrations of individual K^+ -channel blockers were: glibenclamide (Gliben; blocker of ATP-sensitive K^+ channels), 100 nM; apamin (blocker of small conductance Ca^{2+} -activated K^+ channels), 100 nM; charybdotoxin (CTX; blocker of high-conductance Ca^{2+} -activated K^+ channel), 30 nM. * $P < 0.05$ (ANOVA with Scheffe's post-hoc test). *b*, Single-channel openings were recorded in the outside-out patch configuration by depolarizing step pulses from 0 to 50 mV. The concentration of sAPP751 was 1 nM and that of CTX was 30 nM. O, C and P_o indicate open state, close state and open probability, respectively. *c*, Single-channel recordings in an outside-out patch prior to treatment (top), during exposure to 1 nM sAPP751 (middle), and during exposure to sAPP in the presence of 100 nM glibenclamide (bottom).

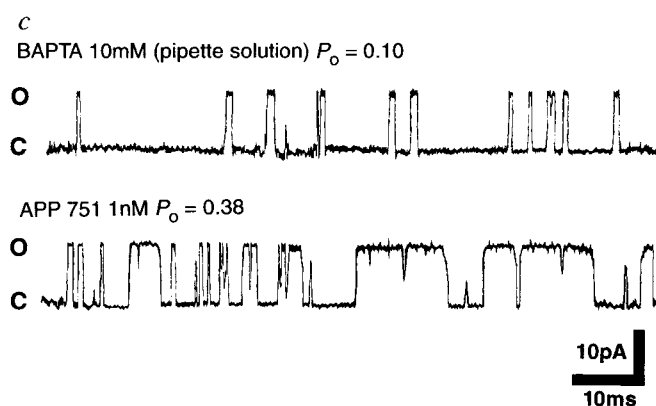
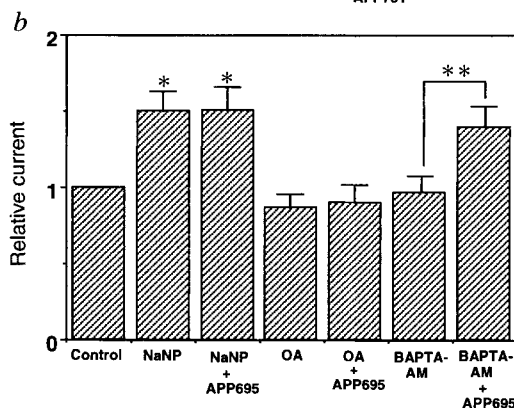
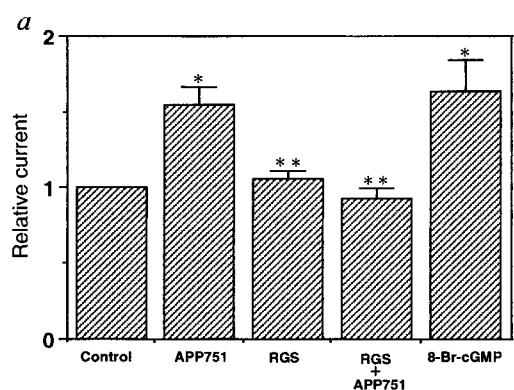
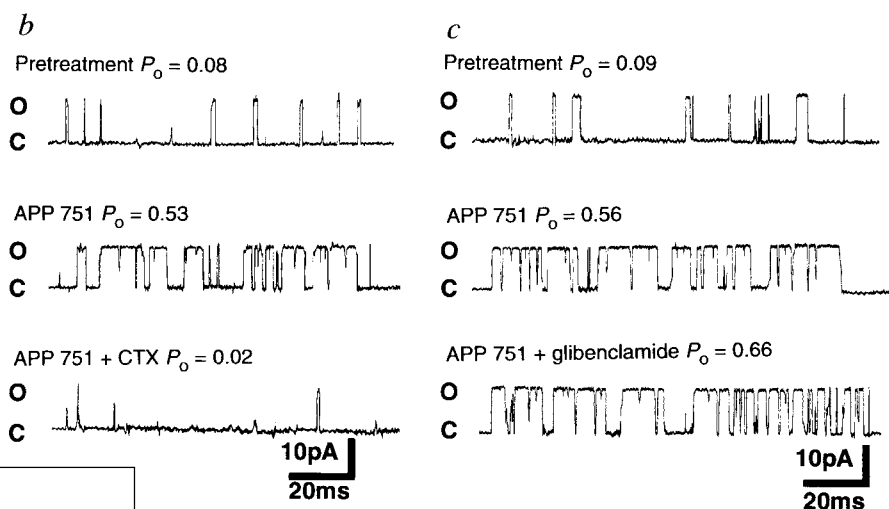
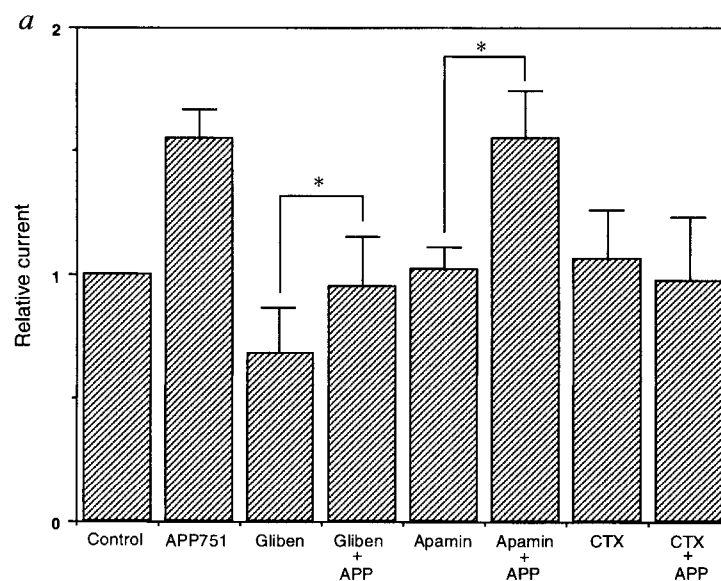


FIG. 3 Evidence that activation of K^+ channels by sAPPs is mediated by cGMP, involves protein phosphorylation, and is Ca^{2+} -independent. *a*, Currents were recorded 2 min after exposure to the treatments indicated. sAPP751, 1 nM; RGS, 30 μ M; 8-bromo-cGMP, 300 μ M. Values are expressed relative to the control value (1.0) and represent the mean and s.e.m. ($n = 5-8$). * $P < 0.01$ compared to control value; ** $P < 0.01$ compared to APP751 value. *b*, Currents were recorded 2 min following exposure to the treatments indicated: sAPP695, 1 nM; sodium nitroprusside (NaNP), 10 μ M; okadaic acid (OA), 10 nM; BAPTA-AM (acetoxymethyl ester), 5 μ M. Values are expressed relative to the control value (1.0) and represent means and s.e.m. ($n = 5-8$). * $P < 0.01$ compared to control value; ** $P < 0.01$. *c*, Single-channel recordings in an outside-out patch with 10 mM BAPTA in the patch pipette. Recordings of single-channel activity before treatment (upper) and during exposure to 1 nM sAPP751 are shown.

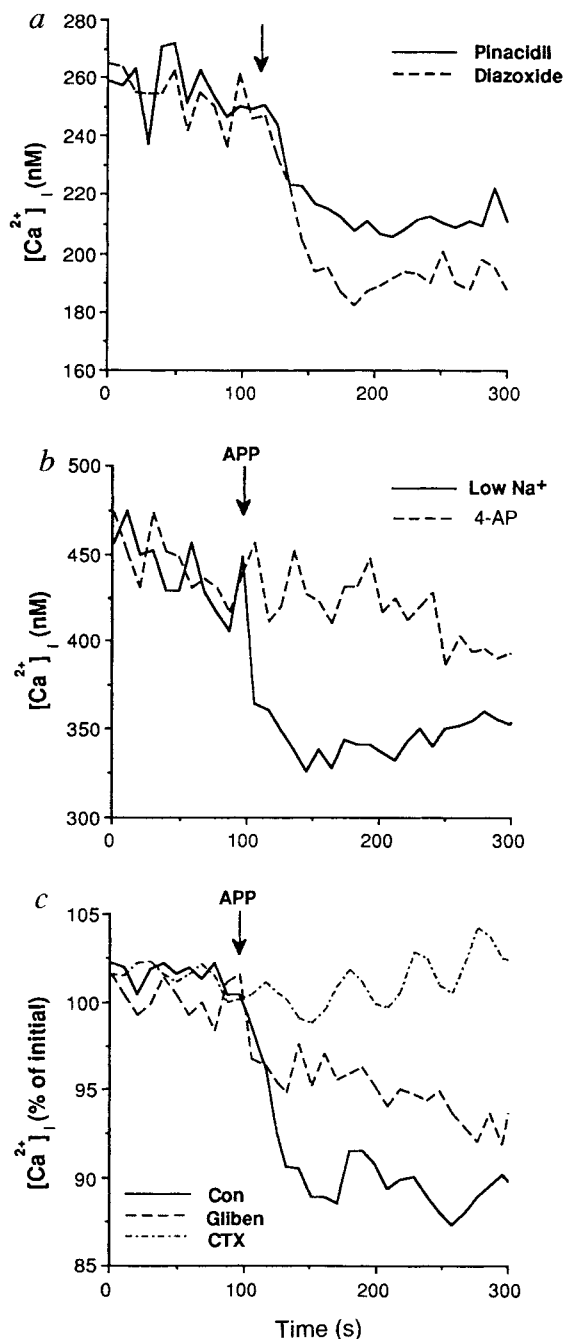


FIG. 4 Role of K^+ channel activity in regulation of neuronal $[Ca^{2+}]_i$. Cultured hippocampal neurons were loaded with Fura-2 and $[Ca^{2+}]_i$ was monitored through fluorescent ratio imaging. *a*, Pinacidil (solid line) or diazoxide (dashed line) was added to $100 \mu M$ at the time indicated by the arrow. *b*, Cultures were either changed to a low-sodium buffer¹¹ (solid line) or pretreated with 5 mM 4-AP for 20 min (dashed line). 200 pM sAPP695 was added at the time indicated by the arrow. *c*, Cultures were pretreated with 10 nM glibenclamide or 30 nM charybdotoxin for 10 min, then compared with untreated cultures for responses to 200 pM sAPP695 (added at the time indicated by the arrow). Each trace represents the mean of 13 cells (*a*) or 26–43 cells (*b* and *c*) from 2–3 plates. **METHODS.** The free $[Ca^{2+}]_i$ in neuronal somata was quantified by fluorescent ratio imaging of the Ca^{2+} indicator dye Fura-2 (ref. 28). Briefly, cells were incubated in the presence of 2 μM Fura-2/AM (acetoxymethyl ester) for 30 min, washed twice (2 ml per wash) with fresh medium, and allowed to incubate for at least 30 min before imaging. Just before imaging, the normal culture medium was replaced with Hank's balanced salt solution (Gibco) containing 10 mM HEPES buffer and 10 mM glucose; cells were then imaged using a Zeiss Atofluor system. The ratio of the fluorescence emission using two different excitation wavelengths (334 nm and 380 nm) was used to determine $[Ca^{2+}]_i$.

is similar to that of 'maxi- K^+ ' channels^{12–14}.

Modulation of ion channels by agonists that induce cGMP production have been reported in a diversity of cell types^{15–18}. Rp-8-Br-cGMPS (RGS), a selective inhibitor of cGMP-dependent protein kinase¹⁹, blocked the activation of I_K by sAPPs (Fig. 3a); RGS alone did not affect I_K . A membrane-permeant cGMP analogue mimicked the activation of I_K by sAPPs (Fig. 3a), an effect that was blocked by CTX (data not shown). Sodium nitroprusside, which induces nitric oxide formation and activation of soluble guanylyl cyclase, also activated the K^+ current, an effect that was not further increased by sAPPs (Fig. 3b). Inhibitors of nitric oxide production, including methylene blue and *N*-methyl-arginine, did not block the $[Ca^{2+}]_i$ -lowering actions of sAPPs¹⁰ consistent with the activation of a membrane-associated guanylyl cyclase²⁰.

Okadaic acid prevented activation of I_K by sAPPs, indicating the involvement of protein phosphatases (Fig. 3b). The intracellular Ca^{2+} -chelating agent BAPTA-AM did not affect the ability of sAPPs to activate I_K (Fig. 3b), and sAPP increased K^+ channel activity in single-channel recordings (outside-out configuration) using an internal solution containing BAPTA (Fig. 3c). These data suggest that sAPPs modulate Ca^{2+} -activated K^+ channels by a Ca^{2+} -independent mechanism involving cGMP kinase and phosphatases^{21,22}. To clarify the relationship between activation of I_K and sAPP effects on neuronal $[Ca^{2+}]_i$ homeostasis^{4,11}, we did imaging studies using the $[Ca^{2+}]_i$ indicator dye Fura-2. We found that the K^+ -channel agonists pinacidil and diazoxide mimicked the ability of sAPPs to depress resting $[Ca^{2+}]_i$ (Fig. 4a). Exposure of neurons to 4-aminopyridine (4-AP) resulted in a modest elevation of $[Ca^{2+}]_i$. sAPP was unable to depress $[Ca^{2+}]_i$ under these conditions, whereas increasing the $[Ca^{2+}]_i$ to an equivalent level by removing Na^+ did not hinder the activity of sAPP (Fig. 4b). sAPPs did not reduce the resting $[Ca^{2+}]_i$ in neurons pretreated with CTX (Fig. 4c), indicating that activation of K^+ channels was required for depression of $[Ca^{2+}]_i$ by sAPPs.

Neuronal activity regulates growth-cone behaviour²³, synaptogenesis²⁴ and synaptic transmission²⁵. Because β -APPs are widely expressed in brain¹, sAPPs are released in response to activity² and neurons from many different brain regions (such as the hippocampus, neocortex or septal area) respond to sAPPs³, our results indicate that sAPPs are probably important for modulating activity-dependent processes throughout the brain. Activation of K^+ channels, membrane hyperpolarization and suppression of action potentials appear to constitute the mechanism whereby sAPPs suppress $[Ca^{2+}]_i$ responses to glutamate and modulate the effects of glutamate on neurite outgrowth and cell survival^{4,5}. Because K^+ channels play key roles in learning and memory processes²⁵, the ability of sAPPs to activate K^+ channels may explain the effects of β -APP on synaptogenesis⁶ and synaptic plasticity⁷. Finally, the present findings suggest that abnormalities in β -APP processing or sAPP activity could explain aberrant K^+ channel properties observed in Alzheimer's disease²⁶ and that these might contribute to the neurodegenerative process by increasing neuronal vulnerability to excitotoxicity and amyloid β -peptide toxicity^{4,27}. □

Received 7 September; accepted 16 October 1995.

- Selkoe, D. J. *Trends Neurosci.* **16**, 403–409 (1993).
- Nitsch, R. M., Farber, S. A., Growdon, J. H. & Wurtman, R. J. *Proc. natn. Acad. Sci. U.S.A.* **90**, 5191–5193 (1993).
- Mattson, M. P. et al. *Trends Neurosci.* **16**, 409–415 (1993).
- Mattson, M. P. et al. *Neuron* **10**, 243–254 (1993).
- Mattson, M. P. *J. Neurobiol.* **25**, 439–450 (1994).
- Roch, J. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **91**, 7450–7454 (1994).
- Huber, S., Martin, J. R., Löffler, J. & Moreau, J.-L. *Brain Res.* **603**, 348–352 (1993).
- Luo, L., Tully, T. & White, K. *Neuron* **9**, 595–605 (1992).
- Mucke, L. et al. *Brain Res.* **666**, 151–167 (1994).
- Barger, S. W., Fiscus, R. R., Ruth, P., Hofmann, F. & Mattson, M. P. *J. Neurochem.* **64**, 2087–2096 (1995).
- Levine, E. S., Dreyfus, C. F., Black, I. B. & Plummer, M. R. *Proc. natn. Acad. Sci. U.S.A.* **92**, 8074–8077 (1995).
- Butler, A., Tsunoda, S., McCobb, D. P., Wei, A. & Salkoff, L. *Science* **261**, 221–224 (1993).
- Pongs, O. *Trends pharmac. Sci.* **13**, 359–365 (1992).
- Wann, K. T. & Richards, C. D. *Eur. J. Neurosci.* **6**, 607–617 (1994).
- Firestein, S. & Zufall, F. *Semin. Cell Biol.* **5**, 39–46 (1994).

16. Sperelakis, N., Tohse, N., Ohya, Y. & Masuda, H. *Adv. Pharmac.* **26**, 217–252 (1994).
17. White, R. E. *et al. Nature* **361**, 263–266 (1993).
18. Archer, S. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **91**, 7583–7587 (1994).
19. Butt, E. *et al. FEBS Lett.* **263**, 47–50 (1990).
20. Barger, S. W. & Mattson, M. P. *Biochem. J.* **311**, 45–47 (1995).
21. Critz, D. & Byrne, J. H. *J. Neurophysiol.* **68**, 1079–1086 (1992).
22. Egan, T. M., Dagan, D. & Levitan, I. B. *J. Neurophysiol.* **69**, 1433–1442 (1993).
23. Kater, S. B., Mattson, M. P., Cohan, C. S. & Connor, J. A. *Trends Neurosci.* **11**, 315–321 (1988).
24. Haydon, P. G. & Drapeau, P. *Trends Neurosci.* **18**, 196–201 (1995).
25. Alkon, D. L. *et al. Brain Res. Rev.* **16**, 193–220 (1991).
26. Etcheberrigaray, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **90**, 8209–8213 (1993).
27. Mark, R. J., Hensley, K., Butterfield, D. A. & Mattson, M. P. *J. Neurosci.* **15**, 6239–6249 (1995).
28. Mattson, M. P., Barger, S. W., Begley, J. G. & Mark, R. J. *Meth. Cell Biol.* **46**, 187–216 (1995).
29. Wakamori, M., Hidaka, H. & Akaike, N. *J. Physiol.* **463**, 585–604 (1993).

ACKNOWLEDGEMENTS. We thank D. Alkon, D. McMahon, R. E. Rydel and S. Younkin for critical comments on an earlier draft of the manuscript, P. Landfield and M. Wakamori for valuable input, and Athena Neurosciences for providing APP antibodies. This research was supported by the NIH, the Alzheimer's Association, and the French Foundation for Alzheimer's Research.

Regulation of glutamate release by presynaptic kainate receptors in the hippocampus

Ramesh Chittajallu, Michel Vignes, Kumlesh K. Dev, Janine M. Barnes*, Graham L. Collingridge & Jeremy M. Henley†

Department of Anatomy, School of Medical Sciences, University of Bristol, University Walk, Bristol BS8 1TD, UK

* Department of Pharmacology, The Medical School, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Most reported actions of kainate are mediated by AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazolepropionate) receptors^{1,2}. Here we report that, unlike AMPA which stimulates, kainate elicits a dose-dependent decrease in L-glutamate release from rat hippocampal synaptosomes and also depresses glutamatergic synaptic transmission. Brief exposure to kainate inhibited Ca^{2+} -dependent [^3H]L-glutamate release by up to 80%. Inhibition was reversed by kainate antagonists but not by the AMPA-selective non-competitive antagonist 1-(4-aminophenyl)-4-methyl-7,8-methylenedioxy-5H-2,3-benzodiazepine (GYKI 52466)^{3–7}. A corresponding reversible kainate-evoked depression of NMDA (*N*-methyl-D-aspartate) receptor-mediated excitatory postsynaptic currents (e.p.s.cs) was observed when AMPA receptors were blocked by GYKI 52466. The synaptic depression was preceded by a brief period of enhanced release and a small inward current was also observed. The effects of kainate were unaffected by metabotropic glutamate (mGlu), GABA_A, GABA_B, glycine and adenosine receptor antagonists. These results indicate that glutamate release can be modulated directly by kainate autoreceptors.

Kainate (100 μM , 2 min) had no effect on uptake or basal release but inhibited both 50 μM 4-aminopyridine (4-AP)- and 25 mM K^{+} -stimulated [^3H]L-glutamate release from hippocampal synaptosomes (Fig. 1a). The action of kainate to inhibit, and that of AMPA to stimulate⁸, [^3H]L-glutamate release was not significantly affected by 100 nM tetrodotoxin (Fig. 1b). Under the conditions used (brief exposure to kainate and static incubation) kainate had a half-maximal inhibitory concentration (IC_{50}) value of 27 μM with inhibition of about 80% at the maximum concentration used (1 mM) (Fig. 1c).

The inhibition of [^3H]L-glutamate release by 100 μM kainate was antagonized by 10 μM 5-nitro-6,7,8,9-tetrahydrobenzo(g)indole-2,3-dione-3-oxime (NS-102), a kainate-selective antagonist^{9,10}. In addition, 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX), which is an antagonist at both AMPA and kainate receptors¹¹, also weakly inhibited the effects of kainate. In contrast, 2,3-dihydroxy-6-nitro-

7-sulphamoyl-benzo(*F*)quinoxaline (NBQX), which is more potent at AMPA than at kainate receptors¹², was ineffective. Neither GYKI 52466, a selective non-competitive AMPA antagonist^{3–7}, nor cyclothiazide, which blocks desensitization at AMPA but not kainate receptors¹³, had any effect on kainate-induced inhibition of K^{+} -stimulated [^3H]L-glutamate release (Fig. 2a). None of the compounds had any effect on basal or K^{+} -stimulated release in the absence of kainate (data not shown), suggesting the absence of any tonic effects. The effects of kainate were unlikely to be mediated indirectly because antagonists of mGlu, GABA_A, GABA_B, glycine and adenosine A1 receptors were all without effect on the kainate-induced inhibition of K^{+} -stimulated [^3H]L-glutamate release (Fig. 2b).

The present findings suggest that the ability of kainate to inhibit the release of acidic amino acids^{14,15} is through an action on kainate rather than AMPA receptors. In order to relate the inhibition of K^{+} -stimulated [^3H]L-glutamate release from synaptosomes to physiological events occurring at hippocampal synapses, we investigated the effects of kainate on synaptic transmission between Schaffer collateral–commissural terminals and CA1 neurons. All experiments included 100 μM GYKI 52466 to block any kainate action on AMPA receptors; NMDA receptor-mediated e.p.s.cs therefore provided a real-time monitor of synaptic glutamate release. In the first series of experiments, we applied 27 μM kainate (the IC_{50} concentration of kainate determined from the synaptosomal experiments) for 2 min. This had four effects. Initially there was a small inward current (38–100 pA; $n = 6$). During the rising phase of this current there was invariably the appearance of an asynchronous burst of NMDA receptor-mediated e.p.s.cs and, in 4 of 6 neurons, an associated short-lived increase in the size of the evoked e.p.s.c. This was rapidly followed by a pronounced depression of the e.p.s.c. (peak effect of $64 \pm 3\%$; $n = 6$; Fig. 3). The recovery from the depression of the e.p.s.c. was slow and sometimes incomplete, whereas the inward current was readily reversible. (The time for 50% recovery from peak effect was 8.0 ± 0.5 and 2.0 ± 0.3 min, respectively ($n = 6$)). Consistent with the synaptosomal release data, neither GABA_B nor adenosine A1 receptor antagonists had any effect on the kainate-evoked effects ($n = 3$; not illustrated).

The sustained depression of the e.p.s.c. was assumed to be the result of a presynaptic mechanism because kainate is not an NMDA receptor antagonist¹⁶. The effect cannot be due to a postsynaptic depolarization¹⁷, which would occur if the clamp was not efficient, because the membrane potential was deliberately held so that the e.p.s.c. was recorded in the region of negative slope conductance (–40 to –45 mV). Thus, any postsynaptic depolarization would result in an increase rather than a decrease in the size of the e.p.s.c. Furthermore, the depression of the e.p.s.c. outlasted the inward current by several minutes. Finally, brief applications of kainate have also been shown to produce long-lasting depressions of AMPA receptor-mediated synaptic transmission in the CA1 region of the hippocampus¹⁸, a result consistent with depression of glutamate release. A possible postsynaptic mechanism which could account for the depression of the e.p.s.c. is calcium entry into the postsynaptic cell which can lead to inactivation of NMDA receptors¹⁹. However, this effect is unlikely to have contributed significantly because similar effects of kainate were also recorded in four neurons which had been extensively dialysed (for between 30 and 97 min) with BAPTA (Fig. 3d).

To obtain a more accurate estimate of potency, agonists were applied for 5–10 min, to allow measurements at, or near, steady state. Under these conditions, the transient appearance of asynchronous e.p.s.cs, the brief increase and ensuing depression of the NMDA receptor-mediated e.p.s.c. (peak depression $52 \pm 3\%$; $n = 17$) and the inward current (10–30 pA) were all observed with 1 μM kainate. Similar effects were also observed with 100 nM domoate ($n = 3$) whereas 10 μM AMPA was ineffective ($n = 3$; Fig. 4a). At a concentration of 300 nM, kainate caused just a sustained enhancement in the e.p.s.c. ($n = 4$; Fig. 4b) whereas at 100 nM it was ineffective ($n = 5$; data not shown).

† To whom correspondence should be addressed.

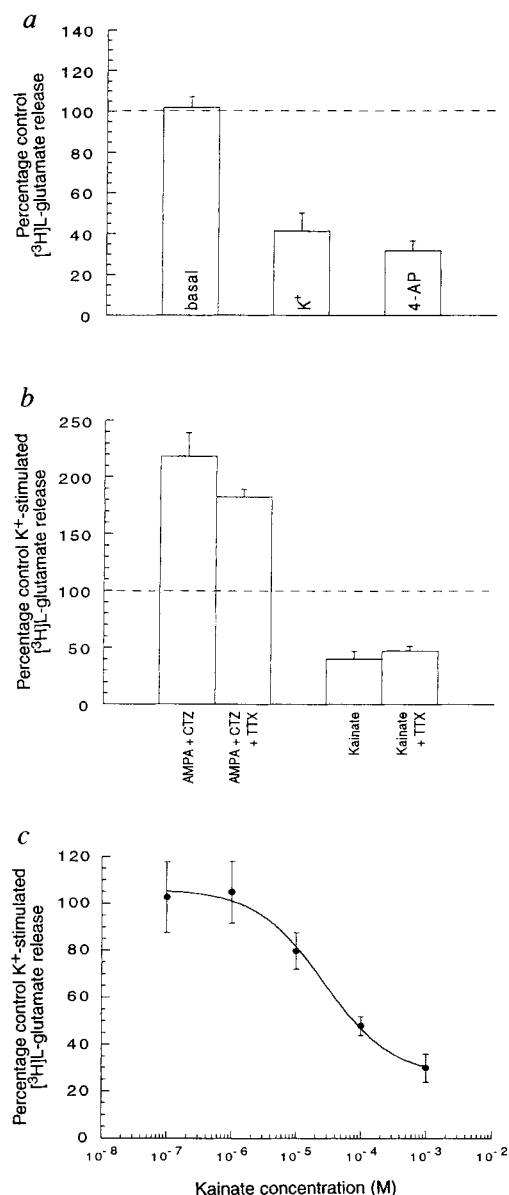


FIG. 1 Kainate inhibits $[^3\text{H}]\text{L-glutamate}$ release from hippocampal synaptosomes. **a**, Effects of $100\text{ }\mu\text{M}$ kainate on basal, 25 mM K^+ -stimulated and $50\text{ }\mu\text{M}$ 4-AP-stimulated $[^3\text{H}]\text{L-glutamate}$ release. **b**, Tetrodotoxin (TTX; 100 nM) does not affect either AMPA ($10\text{ }\mu\text{M}$) potentiation, in the presence of $100\text{ }\mu\text{M}$ cyclothiazide (CTZ), or kainate ($100\text{ }\mu\text{M}$) inhibition of 25 mM K^+ -stimulated $[^3\text{H}]\text{L-glutamate}$ release. **c**, Dose-response curve for kainate-evoked decrease in 25 mM K^+ -stimulated $[^3\text{H}]\text{L-glutamate}$ release. In all cases, the data are the means $\pm$ s.e.m. of at least 3 separate experiments performed in triplicate, each using synaptosomes prepared from 4 hippocampi.

METHODS. Hippocampi were dissected from female Wistar rats ($100\text{--}170\text{ g}$) and synaptosomes prepared as described previously⁸. The synaptosomes were incubated in pre-gassed (O_2/CO_2 ; 95/5%) Krebs buffer (in mM: NaCl 118.0; KCl 4.75; KH_2PO_4 1.2; MgSO_4 1.2; CaCl_2 2.5; NaHCO_3 25.0; glucose 11.0) plus $[^3\text{H}]\text{L-glutamate}$ (25 nM ; 50 Ci mmol^{-1}) at 37°C for 5 min. The uptake reaction was terminated by two washes in 10 vol. ice-cold Krebs buffer and the washed $[^3\text{H}]\text{L-glutamate}$ -loaded synaptosomes were resuspended to 4 ml in fresh ice-cold Krebs, stored on ice and used within 30 min. Electron micrographs of the synaptosomal preparations show structures identical to those reported previously and HPLC analysis confirmed that essentially all the tritium released was in the form of $[^3\text{H}]\text{L-glutamate}$. Greater than 75% of the KCl-stimulated $[^3\text{H}]\text{L-glutamate}$ release was Ca^{2+} -dependent. Aliquots of loaded synaptosomes (0.1 ml) containing $80,000\text{--}100,000\text{ d.p.m.}$ were added to 0.9 ml of either pre-warmed (1) Krebs buffer, (2) high KCl-Krebs buffer (containing an additional 20 mM KCl, which gives roughly half-maximal stimulation of release) or (3) Krebs buffer containing $50\text{ }\mu\text{M}$ 4-AP in the absence (control) or presence of drugs and incubated at 37°C for 2 min. The released fraction of $[^3\text{H}]\text{L-glutamate}$ was separated from the synaptosomal fraction by rapid filtration through GF/B filters followed by a 1 ml wash with ice-cold Krebs buffer. The data are expressed as a percentage of control-stimulated release as defined by the difference between basal and KCl-stimulated $[^3\text{H}]\text{L-glutamate}$ release in the absence of drugs (100%).

FIG. 2 Pharmacology of the kainate-induced inhibition of KCl-stimulated $[^3\text{H}]\text{L-glutamate}$ release. **a**, Antagonism of the effects of kainate ($100\text{ }\mu\text{M}$) by the kainate-selective antagonist NS-102 ($10\text{ }\mu\text{M}$) and by the AMPA/kainate antagonist CNQX ($30\text{ }\mu\text{M}$) but absence of effects of the AMPA receptor-specific potentiating agent cyclothiazide ($100\text{ }\mu\text{M}$), the AMPA receptor-specific non-competitive antagonist GYKI 52466 ($100\text{ }\mu\text{M}$) and the AMPA receptor-selective competitive antagonist NBQX ($10\text{ }\mu\text{M}$). **b**, Kainate-induced inhibition is not reversed by inhibitors of GABA_A receptors (picrotoxin; $50\text{ }\mu\text{M}$), GABA_B receptors (CGP 55845A; $10\text{ }\mu\text{M}$), glycine receptors (strychnine; $100\text{ }\mu\text{M}$), adenosine A1 receptors (8-cyclopentyl-1,3-dipropylxanthine (DPCPX); $0.1\text{ }\mu\text{M}$), group I and II mGluRs ((+)- α -methyl-4-carboxyphenylglycine (MCPG); 1 mM) or group III (L-AP4-sensitive) mGluRs ((S)-2-methyl-2-amino-4-phosphonobutanoate (MAP4); 1 mM). The histograms present the means $\pm$ s.e.m. of the kainate-induced reduction of K^+ -stimulated $[^3\text{H}]\text{L-glutamate}$ release. Each experiment was performed in triplicate using synaptosomes from 4 hippocampi and was replicated on at least 4 separate occasions (the number of times is indicated below the histograms). The first histogram in each figure presents the kainate-induced effect in the absence of other drugs. None of the agents used significantly affected basal release of $[^3\text{H}]\text{L-glutamate}$. * $P < 0.05$; ** $P < 0.01$ (unpaired t -tests).

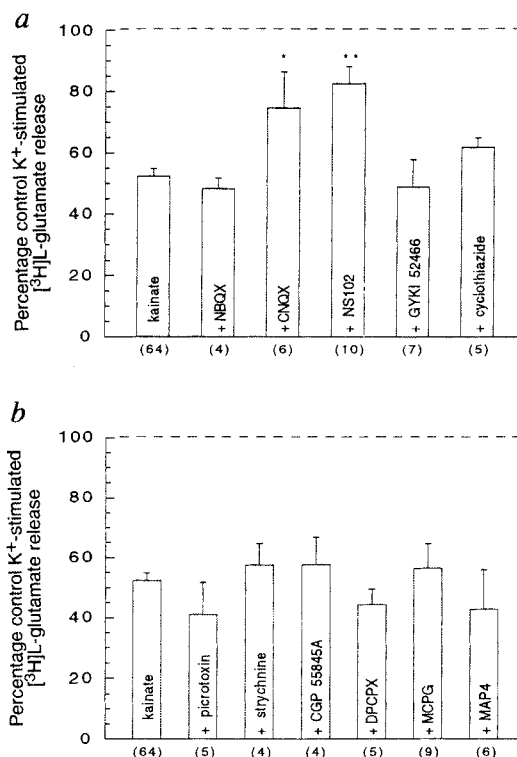
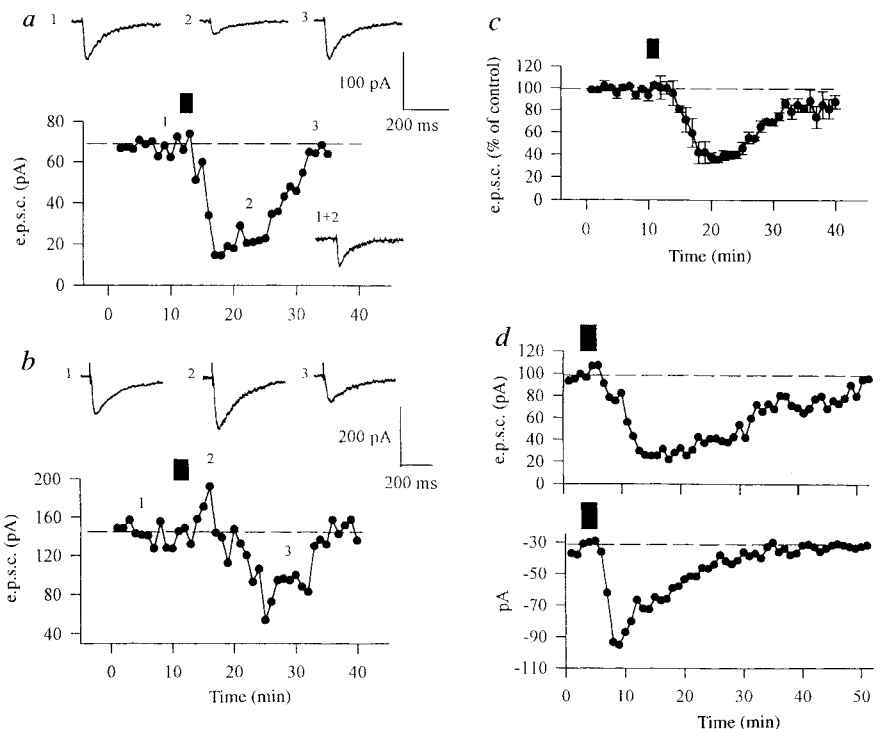


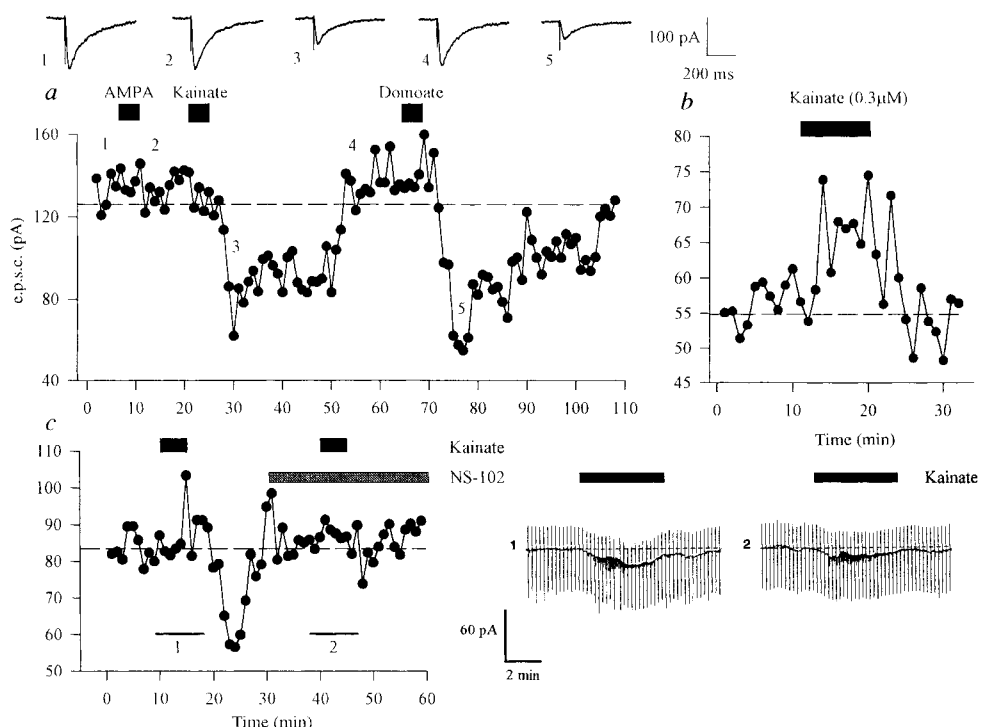
FIG. 3 Kainate depresses synaptic transmission in the hippocampus. *a*, The graph plots the average amplitudes of four successive NMDA receptor-mediated e.p.s.cs versus time. Kainate ($27 \mu\text{M}$) was applied for 2 min at the time indicated by the bar. Representative e.p.s.cs are shown for the times indicated by the numbers. The inset shows a superimposition of traces 1 and 2 (rescaled). *b*, An example from another neuron where the depression was preceded by a transient enhancement of synaptic transmission. *c*, Plots of pooled data for 6 neurons. (The brief increase in e.p.s.c. amplitude following kainate application, seen in 4 of the 6 cells, does not appear in the average because of variations in its time of occurrence). In three of these neurons DPCPX ($0.1 \mu\text{M}$) and CGP 55845 A ($10 \mu\text{M}$) were added to the perfusate to preclude indirect effects mediated through presynaptic adenosine A1 or GABA_B receptors. Because identical results were obtained in the presence and absence of these drugs the data from both sets of experiments have been pooled. *d*, An example of kainate-induced depression of the NMDA receptor-mediated e.p.s.c. (upper trace) and of the inward current (lower trace) in a BAPTA-loaded neuron. Kainate was applied 45 min after obtaining whole-cell access. Note the faster recovery of the inward current than of the synaptic depression, a result typical of both BAPTA-loaded and control neurons.

METHODS. Electrophysiological measurements were carried out on transverse hippocampal slices ($400 \mu\text{m}$) taken from 12–15-day-old Wistar rats using a vibroslice. Slices were collected and maintained in medium comprising (in mM) NaCl 124; KCl 3; NaHCO_3 26; NaH_2PO_4 1.25; CaCl_2 MgSO_4 1; D-glucose 10 (bubbled with O_2/CO_2 , 95/5%). The CA3 region was discarded following dissection of the slices. After at least 1 h equilibration and recovery time, the slices were transferred to a submerged-style recording chamber and perfused with the same medium (at a rate of 2 ml min^{-1}) at room temperature. Whole-cell patch-clamp recordings were obtained from the CA1 region, using the 'blind approach' with glass microelectrodes (5–7 M Ω ; seal resistance $\sim 10 \text{ G}\Omega$) filled with a solution which comprised (in mM) CsMeSO₄ 130; NaCl 1; MgCl_2 1; CaCl_2 0.035; EGTA 0.05; *N*-(2,6-dimethyl-phenylcarbamoylmethyl)-triethylammonium bromide (QX-314) 5; HEPES 5 (adjusted to pH 7.3). For experiments requiring strong calcium chelation, CaCl_2 was omitted, BAPTA (10 mM) added and CsMeSO₄ was reduced (120 mM) to maintain a similar



osmolarity. Microelectrodes were connected to an Axopatch-1D amplifier. The e.p.s.cs were evoked by low-frequency stimulation (one stimulus every 15 s) of the Schaffer collateral–commissural pathway using a bipolar electrode. In all experiments any contribution of GABA_A or postsynaptic GABA_B receptor-mediated conductances was prevented by inclusion of $50 \mu\text{M}$ picrotoxin in the perfusate and Cs⁺ in the patch pipette, respectively. GYKI 52466 ($100 \mu\text{M}$) was present throughout to eliminate the AMPA receptor-mediated component of the e.p.s.c. For all neurons, *I*–*V* curves were constructed and the cell then voltage-clamped at a potential (-40 to -45 mV) which was roughly midway along the region of negative slope conductance. Access resistance was monitored continually and neurons discarded if this parameter changed by more than 20%. Drugs were applied by addition to the perfusing medium. Data are presented as mean $\pm$ s.e.m.

FIG. 4 Pharmacology of the electrophysiological actions of kainate. *a*, Comparison of the effects of 5-min applications of AMPA ($10 \mu\text{M}$), kainate ($1 \mu\text{M}$) and domoate ($0.1 \mu\text{M}$) on NMDA receptor-mediated e.p.s.cs. *b*, The effect of a 10 min application of $0.3 \mu\text{M}$ kainate. *c*, The ability of NS-102 ($10 \mu\text{M}$) to inhibit the effects of kainate ($1 \mu\text{M}$). The right-hand traces are excerpts from the chart record obtained at the times indicated by the corresponding numbers. The large upward deflections are the currents in response to 10-mV voltage commands used to monitor access resistance. The large downward deflections are dominated by the evoked NMDA receptor-mediated e.p.s.cs. The asynchronous downward deflections are NMDA receptor-mediated e.p.s.cs induced by kainate. Note the substantial antagonism of the effects of kainate on the evoked NMDA receptor-mediated e.p.s.cs in this cell and partial antagonism of the kainate-induced inward current. The mean $\pm$ s.e.m. reversal by NS-102 of the kainate-induced synaptic depression and inward current were 55 ± 4 and $33 \pm 6\%$, respectively ($n = 5$).



Repeated applications of 1 μ M kainate induced reproducible effects (up to eight applications tested per neuron). The pre- and, to a lesser extent, postsynaptic effects of 1 μ M kainate were antagonized by 10 μ M NS-102 ($n = 5$) (Fig. 4c) and by 10 μ M CNQX ($n = 3$; data not shown). The NMDA antagonist (D)-2-amino-5-phosphonopentanoate (AP5; 50 μ M) abolished the kainate-induced e.p.s.cs and all evoked synaptic activity but had no effect on the inward current ($n = 3$; data not shown).

Although kainate receptors are widely distributed throughout the central nervous system²⁰ their roles are largely unknown. Previous studies have suggested a presynaptic function on C-fibres in the spinal cord²¹ and on excitatory synapses within the CA3 region of the hippocampus²². The present results using complementary techniques and new pharmacological tools have enabled a direct demonstration that kainate receptors can regulate synaptically released L-glutamate in the hippocampus. The electrophysiological studies revealed a biphasic action of kainate, an initial enhancement of release followed by a long-lasting depression. These results could be explained by a progressive depolarization which would first enhance presynaptic excitability and then, because of inactivation of calcium channels^{23,24}, result in inhibition of neurotransmitter release. Consistent with this interpretation, application of a low dose of kainate induced just the synaptic enhancement. The inward current was an independent effect of kainate which was more readily reversible than the synaptic depression. It is possible that it represents a residual non-desensitizing component of the response described in hippocampal cultures^{9,25,26} or a hippocampal equivalent of the response recorded in the spinal cord²⁷, or certain dorsal root ganglion²⁸ and cerebellar neurons²⁹. The observations that both the pre- and postsynaptic effects were highly sensitive to domoic acid but insensitive to AMPA and were antagonized by NS-102 suggests that the kainate receptors involve, comprise or contain GluR6 subunits³⁰. The presence of both pre- and postsynaptic kainate receptors at CA1 synapses adds to the diversity of glutamate receptors that can influence synaptic function at excitatory synapses. $\square$

Received 7 June; accepted 27 October 1995.

1. Patneau, D. K. & Mayer, M. L. *Neuron* **6**, 785–798 (1991).
2. Bettler, B. & Mulle, C. *Neuropharmacology* **34**, 123–139 (1995).
3. Tarnawa, I., Engeberg, I. & Flatman, J. A. in *Amino Acids, Chemistry, Biology and Medicine* (eds Lubec, G. & Rosenthal, G. A.) 538–546 (ESCOM, Leiden, Netherlands, 1990).
4. Palmer, A. J. & Lodge, D. *Eur. J. Pharmac. molec. Pharm. Sec.* **244**, 193–194 (1993).
5. Donevan, S. D. & Rogawski, M. A. *Neuron* **10**, 51–59 (1993).
6. Zorumski, C. F., Yamada, K. A., Price, M. D. & Olney, J. W. *Neuron* **10**, 61–67 (1993).
7. Paternain, A. V., Morales, M. & Lerma, J. *Neuron* **14**, 185–189 (1995).
8. Barnes, J. M., Dev, K. K. & Henley, J. M. *Br. J. Pharmac.* **113**, 339–341 (1994).
9. Johansen, T. H., Drejer, J., Watjen, F. & Nielsen, E. Ø. *Eur. J. Pharmac. molec. Pharm. Sec.* **246**, 195–204 (1993).
10. Verdoorn, T. A., Johansen, T. H., Drejer, J. & Nielsen, E. Ø. *Eur. J. Pharmac. molec. Pharm. Sec.* **269**, 43–49 (1994).
11. Honoré, T. et al. *Science* **241**, 701–703 (1988).
12. Sheardown, M. J., Nielsen, E. Ø., Hansen, A. J., Jacobsen, P. & Honoré, T. *Science* **247**, 571–574 (1990).
13. Wong, L. A. & Mayer, M. L. *Molec. Pharmac.* **44**, 504–510 (1993).
14. Martin, D., Bustos, G. A., Bowe, M. A., Bray, S. D. & Nadler, J. V. *J. Neurochem.* **56**, 1647–1655 (1991).
15. Zhou, M., Peterson, C. L., Lu, Y.-B. & Nadler, J. V. *J. Neurochem.* **64**, 1556–1566 (1995).
16. Collingridge, G. L. & Watkins, J. C. (eds) *The NMDA Receptor* (Oxford Univ. Press, 1994).
17. Forsythe, I. D. & Clements, J. D. *J. Physiol., Lond.* **429**, 1–16 (1990).
18. Collingridge, G. L., Kehl, S. J. & McLennan, H. J. *Physiol., Lond.* **334**, 33–46 (1983).
19. Legendre, P., Rosenmund, C. & Westbrook, G. L. *J. Neurosci.* **13**, 674–684 (1993).
20. Bahn, S., Volk, B. & Wisden, W. *J. Neurosci.* **14**, 5525–5547 (1994).
21. Agrawal, S. G. & Evans, R. H. *Br. J. Pharmac.* **87**, 345–355 (1986).
22. Sawada, S., Higashimura, M. & Yamamoto, C. *Exp Brain Res.* **60**, 323–329 (1985).
23. Nistri, A. & Cherubini, E. *J. Physiol., Lond.* **435**, 465–481 (1991).
24. Malva, J. O. et al. *Neurosci. Lett.* **185**, 83–86 (1995).
25. Lerma, J., Paternain, A. V., Naranjo, J. R. & Mellström, B. *Proc. natn. Acad. Sci. U.S.A.* **90**, 11688–11692 (1993).
26. Ruano, D., Lambolze, B., Rossier, J., Paternain, A. V. & Lerma, J. *Neuron* **14**, 1009–1017 (1995).
27. Zeman, S. & Lodge, D. *Br. J. Pharmac.* **106**, 367–372 (1992).
28. Huettner, J. E. *Neuron* **5**, 255–266 (1990).
29. Renard, A., Crepel, F. & Audinat, E. *Neuropharmacology* **34**, 335–347 (1995).
30. Egebjerg, J., Bettler, B., Hermans-Borgmeyer, I. & Heinemann, S. *Nature* **351**, 745–748 (1991).

ACKNOWLEDGEMENTS. R.C. and M.V. contributed equally to this work. We thank I. Tarnawa (Institute for Drug Research Ltd, Budapest, Hungary) for GYKI 52466, J. Drejer (NeuroSearch, Glostrup, Denmark) for NS-102, M. Pozza (Ciba Geigy, Basle, Switzerland) for CGP 55845 A and J. C. Watkins (Bristol, UK) for other compounds. This work was supported by the MRC and the Wellcome Trust. K.D. is an MRC-Novo Nordisk collaborative PhD scholar and M.V. is funded by the European Community.

Defective thymocyte proliferation and IL-2 production in transgenic mice expressing a dominant-negative form of CREB

Kevin Barton, Natarajan Muthusamy, Monchai Chanyangam, Christopher Fischer, Cynthia Clendenin & Jeffrey M. Leiden*

Departments of Medicine and Pathology, University of Chicago, Chicago, Illinois 60637, USA

THE basic/leucine zipper (bZip) transcription factor, CREB, binds to the CRE element (TGANNTCA)^{1–5}. The transcriptional activity of CREB requires phosphorylation of the protein on a serine residue at position 119 (ref. 6). CREs are present in a number of T-cell genes^{7,8} but the precise role of CREB in T-cell differentiation and function was unknown. Here we show that resting thymocytes contain predominantly unphosphorylated (inactive) CREB, which is rapidly activated by phosphorylation on Ser 119 following thymocyte activation. T-cell development is normal in transgenic mice that express a dominant-negative form of CREB (CREB_{Al19}, with alanine at position 119) under the control of the T-cell-specific CD2 promoter/enhancer. In contrast, thymocytes and T cells from these animals display a profound proliferative defect characterized by markedly decreased interleukin-2 production, G1 cell-cycle arrest and subsequent apoptotic death in response to a number of different activation signals. This proliferative defect is associated with the markedly reduced induction of *c-jun*, *c-fos*, *Fra-2* and *FosB* following activation of the CREB_{Al19} transgenic thymocytes. We propose that T-cell activation leads to the phosphorylation and activation of CREB, which in turn is required for normal induction of the transcription factor AP1 and subsequent interleukin-2 production and cell-cycle progression.

The total levels of CREB were equivalent in resting and activated normal murine thymocytes (Fig. 1a; α -CREB). Resting thymocytes contained predominantly unphosphorylated CREB as assayed by immunoblotting with an antibody specific for the Ser-119-phosphorylated form of CREB (α -pCREB)⁹ (Fig. 1a). However, thymocyte activation resulted in the phosphorylation of CREB which occurred as early as 1 min after activation and was maximal 5–30 min after activation. Phosphorylation of CREB following thymocyte activation did not require protein kinase A (PKA) because it was not inhibited by preincubation of thymocytes with the PKA inhibitor, H8 (ref. 10) (data not shown). Thus, thymocyte activation results in the rapid phosphorylation and activation of pre-existing CREB through one or more PKA-independent pathways.

To test directly the role of CREB/ATF transcription factors in regulating T-cell development and function, we produced two independent lines of transgenic (Tg) mice that express a dominant-negative form of CREB (CREB_{Al19}) under the transcriptional control of the T-cell-specific CD2 promoter and enhancer^{11,12} (Fig. 1b). This mutant protein inhibits CREB-dependent transcription both *in vitro* and *in vivo*^{13,14}. Thymocytes from the CREB_{Al19} animals expressed high levels of CREB_{Al19} RNA (data not shown) and protein (Fig. 1c). In contrast to mice containing a targeted mutation in the CREB gene¹⁵, overexpression of CREB_{Al19} did not result in compensatory changes in the levels of expression of other CREB/ATF family members including ATF-1, ATF-2, ATF-4 and CREB-2, nor in compensatory increases in phosphorylation of the endogenous CREB protein (Fig. 1c). In fact, significantly lower levels of phosphorylated

*To whom correspondence should be addressed.

CREB and some delay in CREB phosphorylation were observed following activation of the CREB_{A119} thymocytes as compared to NTg control cells (Fig. 1c).

Nuclear extracts from CREB_{A119} thymocytes contained a prominent CRE binding activity (Fig. 1d, arrow) that was present at equivalent levels in both resting and activated thymocytes. This binding activity was significantly greater than the endogenous CRE binding activities in nuclear extracts from control NTg thymocytes (Fig. 1d: compare lanes 6 and 7). This band represented CREB_{A119} binding because (1) it migrated with identical mobility to *in vitro* translated CREB_{A119} homodimers (Fig. 1d, compare lanes 2 and 3); (2) it was specifically competed by

unlabelled CRE-containing oligonucleotides (data not shown); and (3) it was supershifted by an antibody specific for the human CREB protein (Fig. 1d, lane 4). Thus, CREB_{A119} represents the predominant CRE binding activity in the Tg thymocytes, an important consideration in producing a dominant-negative phenotype.

Total numbers of thymocytes and splenocytes were equivalent in the CREB_{A119}Tg and control NTg littermates. Both neonatal and adult Tg thymocytes expressed normal levels of CD3 and TCR α/β , and there were normal numbers and proportions of double-negative (CD4⁻/CD8⁻), double-positive (CD4⁺/CD8⁺)(DP) and single-positive (CD4⁺ and CD8⁺)(SP) thymocytes and splenic T

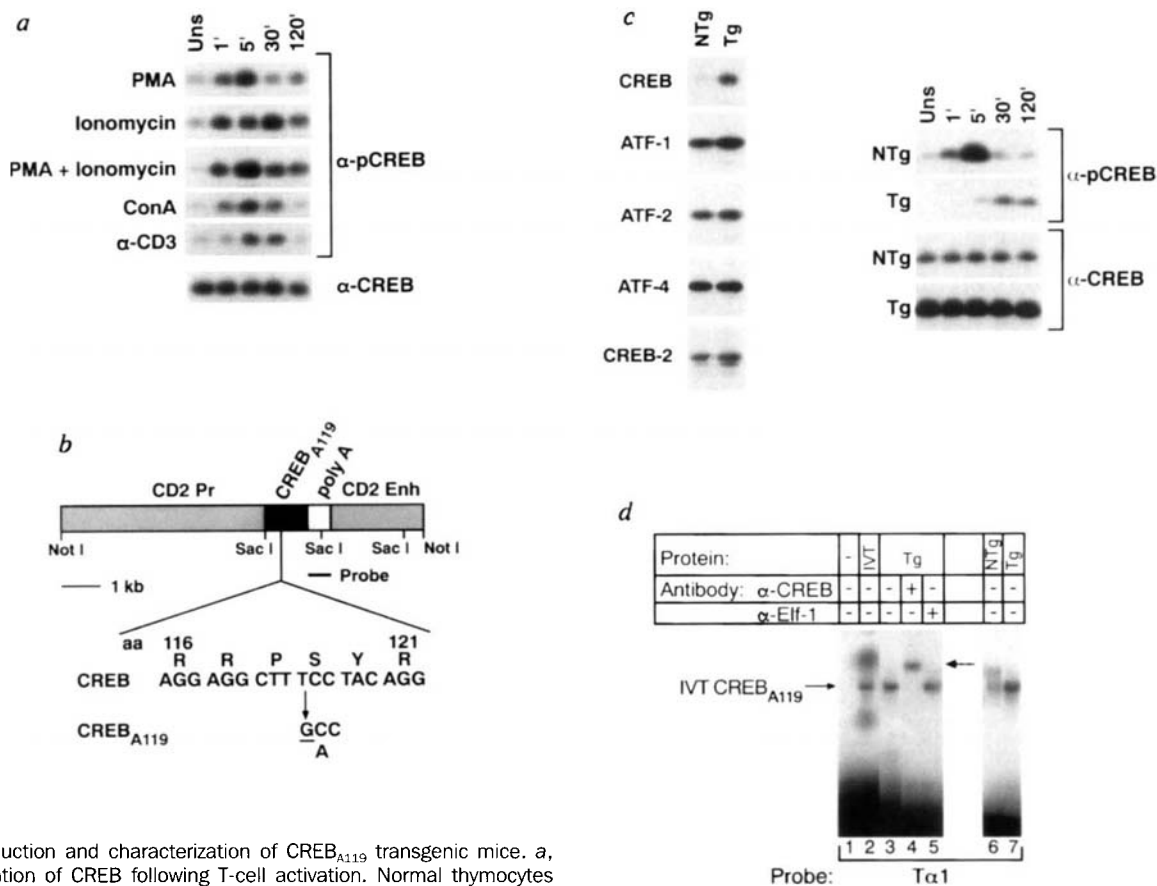


FIG. 1 Production and characterization of CREB_{A119} transgenic mice. **a**, Phosphorylation of CREB following T-cell activation. Normal thymocytes from 4–8-week-old CD1 mice were activated with PMA, ionomycin, PMA + ionomycin, ConA or by immobilized α -CD3 monoclonal antibody (mAb), and cell extracts were assayed by immunoblotting with a polyclonal antibody specific for the Ser-119-phosphorylated form of CREB (α -pCREB) or a mAb that recognizes both phosphorylated and unphosphorylated forms of CREB (α -CREB). Uns, Unstimulated thymocyte extracts. **b**, Schematic representation of the CREB_{A119} transgene. The CD2 promoter (CD2 Pr) and enhancer (CD2 Enh)¹² and the 3' untranslated region and polyadenylation signal from the human CD2 gene (poly A) are shown, as is the probe used in Southern blot analyses (Probe). **c**, Immunoblot analysis of CREB/ATF transcription factors in CREB_{A119} (Tg) and control non-transgenic (NTg) thymocytes. For analysis of CREB phosphorylation (right panels), thymocytes were activated for the indicated times and extracts subjected to immunoblot analysis as described in **a**. No expression of CREM or ATF-3 was observed in either the Tg or NTg thymocytes (data not shown). **d**, Electrophoretic mobility shift assay (EMSA) analysis of CRE binding proteins in nuclear extracts from CREB_{A119} and NTg thymocytes. The binding of *in vitro* translated (IVT) CREB_{A119} is shown with a solid arrow. The position of antibody-supershifted CREB_{A119} is shown by the dotted arrow.

METHODS. Thymocytes (4×10^6 cells per ml) were activated by treatment with PMA (10 ng ml^{-1}), ionomycin ($0.5 \mu\text{g ml}^{-1}$), ConA ($2 \mu\text{g ml}^{-1}$) or immobilized α -CD3 mAb, 145.2C11 ($16 \mu\text{g ml}^{-1}$), for the times shown (in

minutes). Protein samples corresponding to 4×10^6 cells were subjected to immunoblot analysis as described²⁷ using a rabbit polyclonal IgG antibody specific for the Ser 119 phosphorylated form of CREB⁹ (α -pCREB) (1:3,000 dilution) or the mAb, 24H4B (1:50 dilution) (Santa Cruz Biotechnology), which recognizes both phosphorylated and unphosphorylated forms of CREB (α -CREB). Immunoblots containing cell extracts from 2×10^6 thymocytes were probed for ATF-1, ATF-2, ATF-4 and CREB-2 using commercially available monoclonal antibodies, C41-5.1, C-19 and Z5 (which reacts with both ATF-4 and CREB-2), respectively (1:150 dilutions) (Santa Cruz Biotech). The dominant-negative CREB used in these studies was produced by PCR-mediated mutagenesis of the human CREB cDNA¹ at Ser 119 (Ser 119 to Ala) thereby creating a non-phosphorylatable form of CREB (CREB_{A119})²⁷. The CREB_{A119} cDNA was ligated into the XhoI site of the CD2 transgene vector. The 8.7-kb NotI transgene fragment (shown in **b**) was microinjected into the male pronucleus of fertilized single-cell CD1 embryos. Transgenic founders were identified by Southern blot analysis of tail DNA. The preparation of nuclear extracts, and the EMSAs were performed as described previously²⁷. The α -CREB mAb, 24H4B (2 μl ; Santa Cruz Biotechnology) and 2 μl α -E1f-1 mAb, 5A3, were used in the antibody supershift experiments. The CRE-containing T α 1 oligonucleotide was used as a radiolabelled probe in all EMSAs²⁸.

cells in the Tg animals (data not shown). Thus, CREB_{A119} expression did not noticeably disrupt T-cell development. In contrast, the CREB_{A119} thymocytes displayed a profound defect in proliferation following activation with concanavalin A (conA), PMA + ionomycin (P + I), or immobilized α -CD3 monoclonal antibody (Fig. 2a–c). This proliferative defect was observed at 24, 48 and 60 h after stimulation and could not be rescued by increased amounts of PMA or α -CD3. Moreover, it could not be reversed by treatment with an α -CD28 antibody, which can provide a costimulatory signal to T cells (data not shown). Splenic T cells from the CREB_{A119} animals expressed significantly lower levels of CREB_{A119} than thymocytes from the same animals and displayed a significant, but less profound (39%) reduction in proliferation in response to activation by P + I as compared to control NTg T cells ($P < 0.001$) (data not shown). Thus, the magnitude of the T-cell proliferative defect appeared to be proportional to the levels of transgene expression in the CREB_{A119} T cells.

T-cell receptor (TCR) crosslinking leads to the production of soluble interleukin-2 (IL-2) and to the induction of a high-affinity IL-2 receptor (IL-2R) on the surface of activated T cells¹⁶. Subsequent binding of IL-2 to its high-affinity receptor is required for T-cell proliferation¹⁷. IL-2 production was decreased by more than 99% in the CREB_{A119} thymocytes as compared to control NTg thymocytes following activation with P + I (Fig. 2d). Similar results were obtained following α -CD3 activation (data not shown). In contrast, both the CREB_{A119} and NTg thymocytes upregulated expression of the high-affinity IL-2R equivalently (IL-2R α shown in Fig. 2e; IL-2R β and γ , data not shown). Decreased IL-2 production alone was not responsible for the proliferative defect of the CREB_{A119} thymocytes, because even high levels of exogenously supplied IL-2 failed to restore fully the proliferation of CREB_{A119} thymocytes following activation (Fig. 2f). Thus, there must be additional defective proliferative pathways in these cells.

Following activation with P + I, the CREB_{A119} thymocytes displayed an 80% decrease in viability (as assessed by the size of the hypodiploid peak following propidium iodide staining) and G1 cell-cycle arrest as compared to the NTg thymocytes (Fig. 3a, b). Thirty-six hours after activation, fewer than 15% of the viable CREB_{A119} thymocytes had traversed the G1–S checkpoint of the cell cycle, as compared to 60% of the viable NTg cells. The

CREB_{A119} thymocytes displayed similar significantly increased rates of cell death and impaired cell-cycle progression at 18, 24 and 60 h after activation (data not shown).

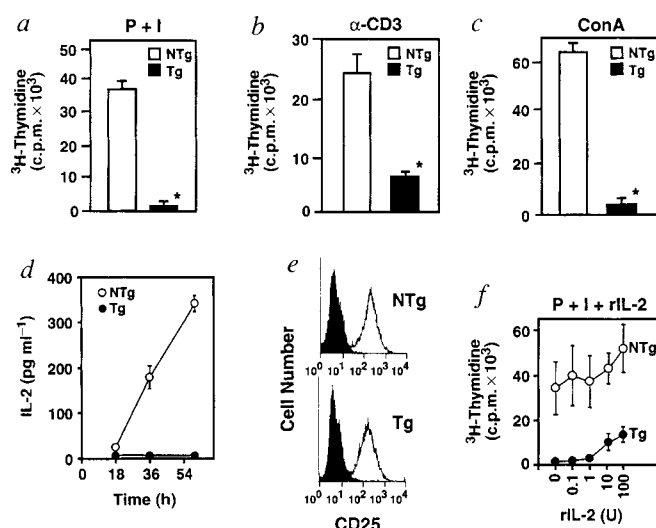
Thymocytes are composed of distinct cell populations which respond differently to activational signals. For example, DP cells tend to undergo apoptosis in response to TCR crosslinking, whereas more mature SP thymocytes proliferate in response to TCR crosslinking¹⁸. To assess the proliferative and apoptotic potentials of these different thymocyte populations, CREB_{A119} thymocytes were activated with conA and stained with 7-amino-actinomycin D (7-AAD) and antibodies specific for CD4 and CD8. The viability of unstimulated CREB_{A119} thymocytes was identical to that of NTg thymocytes during 72-h culture *in vitro* (data not shown). In contrast, 48 h after ConA activation, the viabilities of the SP CREB_{A119} thymocytes were reduced by about 90% as compared to control NTg thymocytes ($P < 0.0002$) (Fig. 3c, d). This decreased viability was associated with significantly increased rates of apoptosis of the activated SP CREB_{A119} thymocytes (Fig. 3c, d). Similar results were obtained at 24 and 72 h after activation (data not shown). The DP Tg thymocytes also demonstrated reduced viabilities at both 24 and 48 h after activation (Fig. 3c, d). Thus, the CREB_{A119} SP thymocytes displayed both a G1 cell-cycle arrest and increased rates of apoptotic cell death in response to multiple activation signals.

AP1 is an important regulator of T-cell activation and IL-2 transcription^{19,20}. Previous studies have suggested that CREB plays an important role in regulating AP1 expression in multiple cell types^{21–23}. The expression of *Fra-1*, *JunB* and *JunD*, were similar or identical in the CREB_{A119} and NTg thymocytes following activation with P + I (Fig. 4). In contrast, the inducible expression of *c-jun*, *c-fos*, *Fra-2* and *FosB* were reduced by 55%, 60%, 79% and 87%, respectively. In control experiments, the patterns of expression of *c-myc* and *GPD* were indistinguishable in Tg and NTg thymocytes (Fig. 4). Thus, the inducible expression of multiple AP1 family members was markedly and specifically decreased in CREB_{A119} thymocytes.

The results described in this report represent the first direct evidence of the importance of CREB/ATF transcription factors in regulating IL-2 production and T-cell proliferation. Taken together, they are consistent with a model in which TCR crosslinking results in CREB phosphorylation (and activation) which,

FIG. 2 A proliferative defect following activation of CREB_{A119} transgenic thymocytes. a–c, Freshly isolated CREB_{A119} (Tg) or control NTg thymocytes were activated with PMA + ionomycin (a), α -CD3 crosslinking (b), or ConA (c) and proliferation was measured by ³H-thymidine incorporation. The results are shown as the mean $\pm$ s.e.m. of 10 (a), 5 (b) or 3 (c) independent experiments using thymocytes from two independently derived CREB_{A119} transgenic lines. d, IL-2 production by activated CREB_{A119} or control NTg thymocytes (NTg). The results are expressed as mean $\pm$ s.e.m. for thymocytes obtained from three animals. e, IL-2 receptor α expression on activated CREB_{A119} (Tg) and control NTg thymocytes. Thymocytes were activated with PMA + ionomycin for 36 h, stained with FITC-labelled α -CD25 mAb (unshaded curve) or an isotype matched control mAb (shaded curve) and analysed by FACS on a Becton Dickinson flow cytometer. The X-axis represents the mean fluorescence intensity. f, Recombinant IL-2 fails to rescue the decreased proliferation of CREB_{A119} thymocytes. CREB_{A119} (Tg) or control NTg thymocytes were activated by treatment for 62 h with PMA + ionomycin. The indicated amounts of recombinant IL-2 were added to each culture at the time of activation. Thymocyte proliferation was measured by ³H-thymidine incorporation. The results are shown as the mean $\pm$ s.e.m. of 3 independent experiments. Asterisks indicate significant differences ($P < 0.0001$) as compared to NTg thymocytes.

METHODS. For measurements of thymocyte proliferation, freshly isolated thymocytes (2.5×10^5) were cultured in triplicate in 200 μ l DMEM medium + 10% FCS (GIBCO-BRL) in 96-well tissue culture plates (Falcon). Thymocytes were activated for 62 h at 37°C by treatment with PMA (10 ng ml⁻¹) and ionomycin (0.5 μ g ml⁻¹), the α -CD3 mAb, 145.2C11 (16 μ g ml⁻¹) immobilized on plastic tissue culture plates, or Con A (2 μ g ml⁻¹). Forty-four h after activation, cells were labelled for 18 h by incubation in ³H-thymidine-containing tissue culture medium (1 μ Ci ml⁻¹; specific activity, 2 Ci mmol⁻¹) (Amersham). Cells were collected onto glass fibre



filter mats using a Filtermate 196 cell collector and ³H-thymidine incorporation was measured in a Topcount microplate scintillation counter (Packard). For measurements of IL-2 production, thymocytes were activated for 36 h with PMA (10 ng ml⁻¹) + ionomycin (0.5 μ g ml⁻¹). Culture supernatants from the activated thymocytes were assayed by ELISA in duplicate for IL-2 activity using commercially available antibodies according to the manufacturer's instructions (Pharmingen).

in turn, is required for AP1 induction. The presence of a functionally important CREB binding site in the *c-fos* promoter^{21,22} suggests that phosphorylated CREB may directly activate *c-fos* transcription. The mechanism underlying the observed decreases in *c-jun*, *Fra-2* and *FosB* induction in the CREB_{A119} thymocytes requires further investigation. AP1 binding to multiple sites in the IL-2 promoter is essential for inducible IL-2 transcription^{19,20}.

Thus, the observed defect in IL-2 production in the CREB_{A119} thymocytes may result, in part, from defective AP1 induction. Finally, activation of T cells in the absence of IL-2 can lead to apoptosis¹⁶ as can forced expression of *c-myc* in the absence of growth factors²⁴⁻²⁶, thereby perhaps explaining the increased rates of apoptosis of the activated CREB_{A119} thymocytes. The defect in AP1 induction (as well as other effects of over-expression of

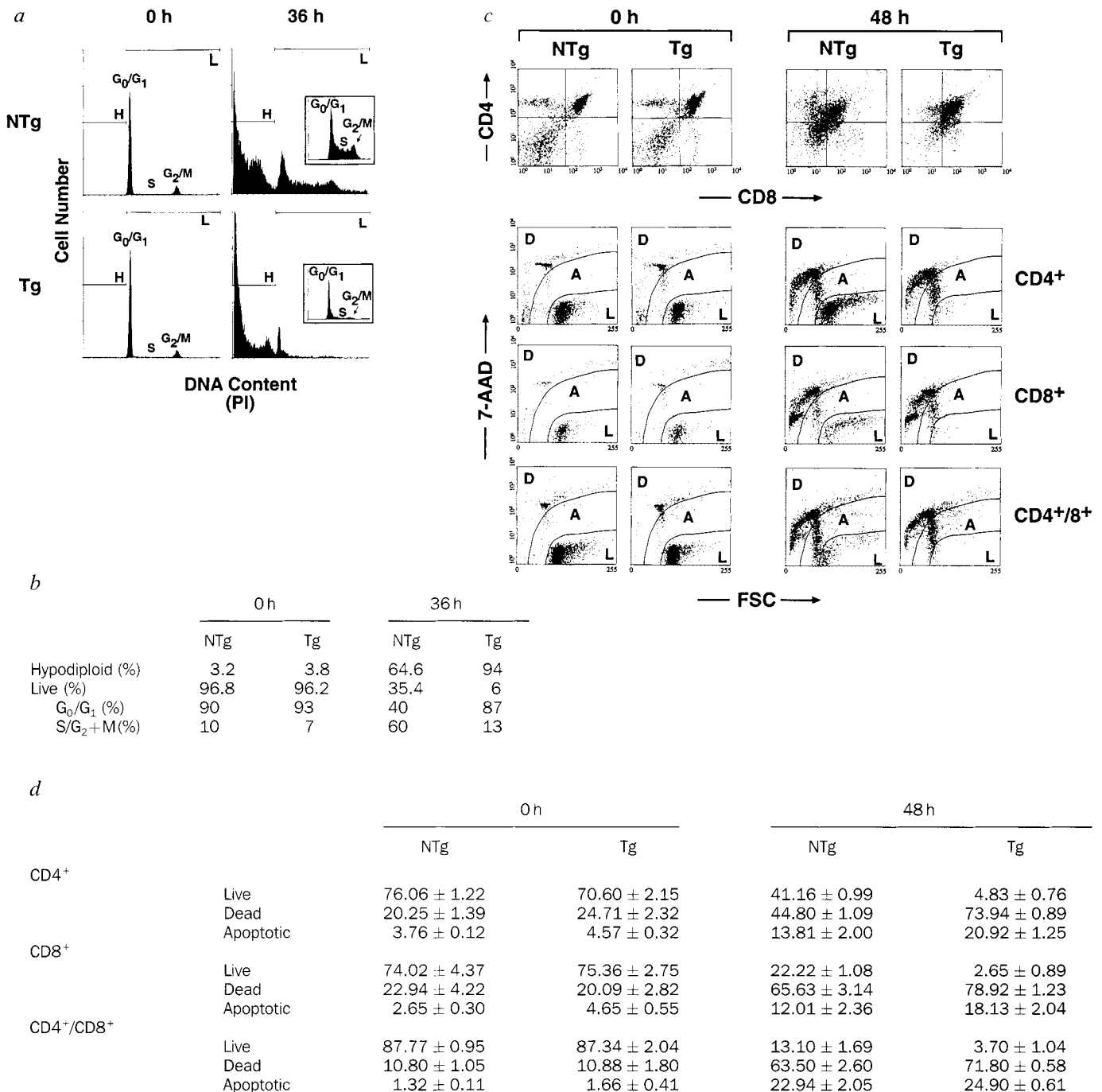


FIG. 3 G1 cell-cycle arrest and increased apoptosis following activation of CREB_{A119} thymocytes. **a, b**, CREB_{A119} (Tg) or control NTg thymocytes were activated for 36 h by treatment with PMA (10 ng ml⁻¹) + ionomycin (0.5 µg ml⁻¹) and stained with propidium iodide (50 µg ml⁻¹). Hypodiploid (H) and live (L) cells were quantified using WINmidi software. Cell-cycle progression was analysed in resting (0h) and activated (36h) viable thymocytes using a Becton Dickinson FACScan and CellFit software (see inserts for cell cycle analysis of 36-h P + I activated thymocytes). **c, d**, CREB_{A119} (Tg) or control (NTg) thymocytes were cultured in medium alone

(0h) or activated for 48 h by treatment with ConA (2 µg ml⁻¹) and stained with 7-amino-actinomycin D (7-AAD) and PE-labelled α-CD4 and FITC-labelled α-CD8 mAbs. Ungated cell populations were analysed by flow cytometry (**c**, top panels). Viable (L), apoptotic (A) and dead (D) cells from individually gated populations of CD4⁺, CD8⁺ and CD4⁺/CD8⁺ thymocytes were identified by low-, intermediate- and high-level 7-AAD staining as described²⁹ (**c**, bottom panels). The results represent the mean ± s.e.m. for three Tg and three NTg animals.

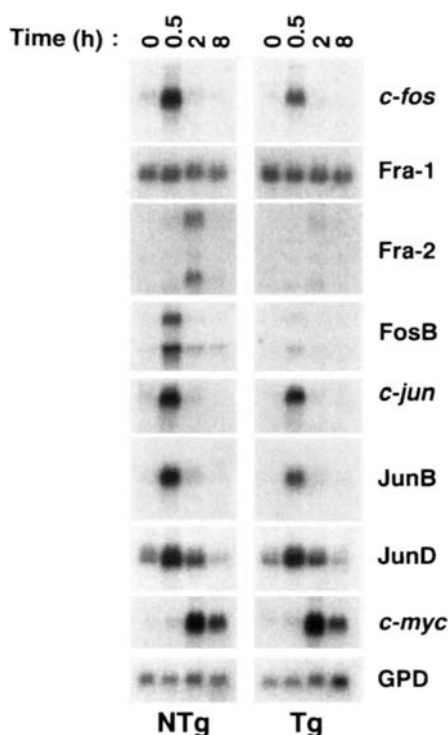


FIG. 4 Northern blot analysis of AP1 expression in activated CREB_{A119} thymocytes. Thymocytes from CREB_{A119} transgenic (Tg) mice or non-transgenic (NTg) control littermates were activated for the times shown with PMA (10 ng ml⁻¹) + ionomycin (0.5 µg ml⁻¹) and RNA was analysed by northern blot analysis with probes specific for different AP1 family members. In control experiments, the same RNA samples were analysed in parallel with probes for *c-myc* and glucose-3-phosphate dehydrogenase (GPD). METHODS. RNA was extracted using guanidinium hydrochloride. Northern blots containing 10 µg of total RNA per lane were prepared and hybridized to the following radiolabelled cDNA probes: murine *c-fos* (a 0.7 kb *Sall*-*PstI* fragment), murine *Fra-1* (the full-length 1.4 kb cDNA), human *Fra-2* (the 1.0 kb full-length cDNA), murine *FosB* (a 1.6 kb *HindIII* fragment), murine *c-jun* (a 1.0 kb *BamHI*-*EcoRI* fragment), murine *junB* (a 0.4 kb *BamHI* fragment), murine *junD* (the full-length cDNA), murine *c-myc* (a 1.9 kb *HindIII* fragment), chicken GPD (the full-length cDNA). Band intensities were quantified on a Molecular Dynamics phosphorimager.

CREB_{A119}) may also perturb other important signalling pathways, as suggested by our finding that the proliferative defect in the CREB_{A119} thymocytes cannot be reversed by exogenous IL-2 alone. Given the importance of CREB in regulating both T-cell proliferation and apoptosis, it will be of interest to understand more fully the signalling pathways that regulate CREB phosphorylation following T-cell activation, and the role of CREB/ATF proteins in modulating the expression of specific cell-cycle regulatory molecules in T cells. The CREB_{A119} mice represent a useful model system for addressing these questions. □

Received 19 July; accepted 20 October 1995.

- Hoeffler, J. P., Meyer, T. E., Yun, Y., Jameson, J. L. & Habener, J. F. *Science* **242**, 1430–1433 (1988).
- Gonzalez, G. A. et al. *Nature* **337**, 749–752 (1989).
- Habener, J. F. *Molec. Endocr.* **4**, 1087–1094 (1990).
- Lee, K. A. W. & Masson, N. *Biochim. biophys. Acta* **1174**, 221–223 (1993).
- Vallejo, M. J. *Neuroendocrinology* **6**, 587–596 (1994).
- Gonzalez, G. A. & Montminy, M. R. *Cell* **59**, 675–680 (1989).
- Leiden, J. M. A. *Rev. Immun.* **11**, 539–570 (1993).
- Calame, K. & Eaton, S. *Adv. Immun.* **43**, 235–275 (1988).
- Ginty, D. D. et al. *Science* **260**, 238–241 (1993).
- Hidaka, H., Inagaki, M., Kawamoto, S. & Sasaki, Y. *Biochemistry* **23**, 5036–5041 (1984).
- Greaves, D. R., Wilson, F. D., Lang, G. & Kioussis, D. *Cell* **56**, 979–986 (1989).
- Lake, R. A., Wotton, D. & Owen, M. J. *EMBO J.* **9**, 3129–3136 (1990).
- Lee, C. Q., Yun, Y., Hoeffler, J. P. & Habener, J. F. *EMBO J.* **9**, 4444–4465 (1990).
- Struthers, R. S., Vale, W. W., Arias, C., Sawchenko, P. E. & Montminy, M. R. *Nature* **350**, 622–624 (1991).

- Hummler, E. et al. *Proc. natn. Acad. Sci. U.S.A.* **91**, 5647–5651 (1994).
- Smith, K. *Curr. Opin. Immun.* **4**, 271–276 (1992).
- Rothenberg, E. V. *Adv. Immun.* **51**, 85–214 (1992).
- Kisielow, P. & von Boehmer, H. *Adv. Immun.* **58**, 87–209 (1995).
- Jain, J., McCaffrey, P. G., Valge, A. V. & Rao, A. *Nature* **356**, 801–804 (1992).
- Boise, L. H. et al. *Molec. cell. Biol.* **13**, 1911–1919 (1993).
- Berkowitz, L. A., Riabowol, K. T. & Gilman, M. Z. *Molec. cell. Biol.* **9**, 4272–4281 (1989).
- Sheng, M., Dougan, S. T., McFadden, G. & Greenberg, M. E. *Molec. cell. Biol.* **8**, 2787–2796 (1988).
- Ofir, R., Dwarki, V. J., Rashid, D. & Verma, I. M. *Gene Express.* **1**, 55–60 (1991).
- Wagner, A. J., Kokontis, J. M. & Hay, N. *Genes Dev.* **8**, 2817–2830 (1994).
- Shi, Y. et al. *Science* **257**, 212–214 (1992).
- Bennett, M. R., Fanidi, A. & Evan, G. I. *EMBO J.* **13**, 3286–3295 (1994).
- Wang, C. Y., Petryniak, B., Thompson, C. B., Kaelin, W. G. & Leiden, J. M. *Science* **260**, 1330–1335 (1993).
- Karpinski, B. A., Morie, G. D., Huggenvik, J., Uhler, M. D. & Leiden, J. M. *Proc. natn. Acad. Sci. U.S.A.* **89**, 4820–4824 (1992).
- Schmid, I., Uittenbogaart, C. H., Keld, B. & Giorgi, J. V. *J. Immun. Meth.* **170**, 145–157 (1994).

ACKNOWLEDGEMENTS. K.B. and N.M. contributed equally to this work. We thank J. Habener for the gift of the human CREB cDNA, M. Owen for generously providing the CD2 transgenic vector and M. Greenberg for the gift of the α -pCREB antibody. We thank M. S. Parmacek, J. Bluestone and M. C. Simon for helpful discussions and critical review of the manuscript. K. Siegrist provided excellent technical assistance, L. R. Gottschalk helped with the preparation of the figures and S. Vander Veer provided help with the preparation of the manuscript. This work was supported in part by an NIH grant to J.M.L.

Sensitization of cells and retroviruses to human serum by (α 1-3) galactosyltransferase

Yasuhiro Takeuchi*, Colin D. Porter*, Karen M. Strahan†, Andrew F. Preece†, Kenth Gustafsson†, Francois-Loic Cosset*, Robin A. Weiss* & Mary K. L. Collins*

* Chester Beatty Laboratories, Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK

† Institute of Child Health, University of London, 30 Guilford Street, London WC1N 1EH, UK

MAMMALIAN C-type retroviruses are inactivated by human serum^{1,2}, following triggering of the classical complement cascade³. This may have inhibited transmission to humans of C-type oncoviruses from other mammals¹. Indeed, the retroviruses human immunodeficiency virus and human T-cell leukaemia virus are resistant to human complement^{4,5}. Antibody-independent activation of human C1q, the first component of the classical pathway, by retroviral envelope proteins has been described⁶. However, retroviruses produced from human cells are resistant to inactivation by human complement^{7,8} and human serum is known to contain antibodies directed against carbohydrates on retroviral envelopes^{9–11}. Gal(α 1-3)Gal terminal carbohydrates are expressed by most mammals but are absent in humans, which lack a functional (α 1-3)galactosyltransferase gene^{12,13}. Here, we demonstrate that anti-Gal(α 1-3)Gal antibodies in human serum inactivate retroviruses produced from animal cells. Expression of porcine (α 1-3)galactosyltransferase^{14–16} in human cells renders the cells and the retroviruses they produce sensitive to human serum.

C-type retroviruses produced from murine or canine cells, but not from human cells, are sensitive to inactivation by human serum, paralleling the sensitivity of the cells to human serum lysis^{7,8}. Figure 1a–c shows that human plasma antibodies bound strongly to murine or canine cells, but only weakly to human cells. A large proportion of the antibodies that bound murine and canine cells were anti-Gal(α 1-3)Gal, as demonstrated by their retention on a Gal(α 1-3)Gal column. The low-level binding of antibodies to human cells may be unspecific to Gal(α 1-3)Gal groups as there was similar level of binding with human plasma depleted of anti-Gal(α 1-3)Gal (Fig. 1c). Furthermore, the murine and canine cells, but not human cells, stained brightly with *Bandeiraea simplicifolia* isolectin B4 (BS-IB4), which specifically recognizes α -galactosyl groups (Fig. 1d–f). BS-IB4

also detected proteins the size of the amphotropic murine leukaemia virus (MLV-A) or feline endogenous RD114 viral gp70 envelope surface glycoprotein (SU) in virions produced by murine or canine but not human cells (Fig. 2b).

We examined whether retrovirus inactivation might be mediated by anti-Gal(α 1-3)Gal binding to viral envelopes, resulting in complement activation. Addition of the disaccharide Gal(α 1-3)galactose, but not the isomer Gal(β 1-4)galactose, inhibited retrovirus inactivation by human serum (Table 1). Furthermore, depletion of anti-cell antibodies by pre-adsorption of human serum on murine or canine cells blocked subsequent inactivation of MLV-A from mouse cells and RD114 virus from dog cells (Table 1). As predicted, serum pre-adsorption on human cells did not affect virus inactivation. Heat treatment of serum to destroy complement likewise abolished virus inactivation, which could be restored by addition of anti-cell antibody-depleted serum (Table 1). Thus, both anti-cell antibodies and complement are required for inactivation of retroviruses by human serum.

To verify the involvement of Gal(α 1-3)Gal epitopes in virus inactivation, human HT1080 cells were transfected with a porcine cDNA encoding (α 1-3)galactosyltransferase [(α 1-3)GT]. A

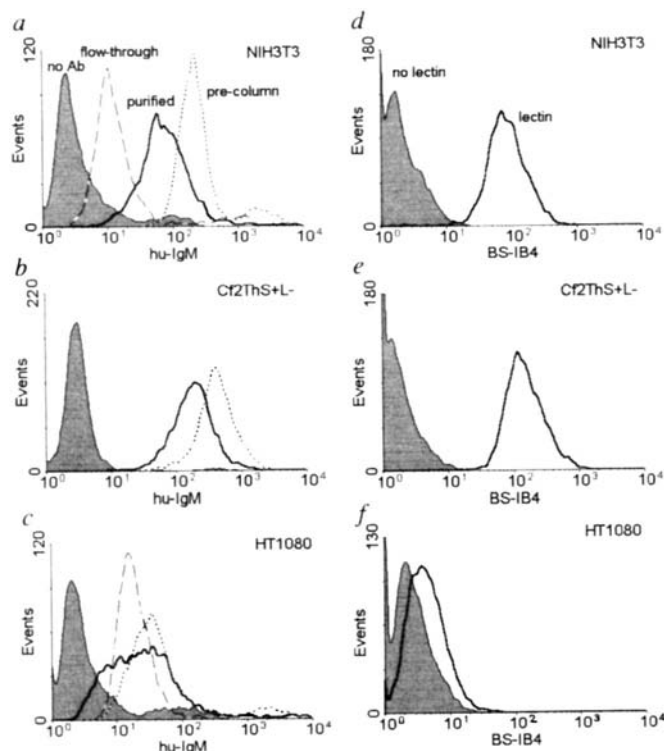


FIG. 1 Expression of Gal(α 1-3)Gal groups at the cell surface. Mouse NIH3T3 (a and d), dog Cf2ThS+L- (b and e) and human HT1080 cells (c and f) were collected using a scraper, washed twice with phosphate-buffered saline (PBS) and then resuspended in PBS supplemented with 1% BSA and 0.1% NaN_3 (PBA) at 1×10^7 cells per ml. a-c, 25 μ l of cell suspension was mixed with 50 μ l of pooled human plasma (dotted line), antibodies purified from this plasma by retention on a Gal(α 1-3)Gal(β 1-4)GlcNAc conjugated column (bold line), flow-through of the column (broken line) (BioTransplant Inc.) or PBA (filled histogram), then incubated at 4°C for 1 h. Cells were washed twice with PBA, resuspended in 50 μ l of FITC-conjugated goat anti-human IgM antibodies (Zymed Inc.), incubated at 4°C for 1 h, washed twice with PBA and then fixed in PBS supplemented with 2% formaldehyde. IgM was detected as representative of anti-Gal(α 1-3)Gal antibodies, as the trisaccharide column appears to retain IgM more efficiently than IgG (KMS, AFP and KG, unpublished). d-f, 25 μ l of cell suspension was mixed with 25 μ l of PBA with or without $10 \mu\text{g ml}^{-1}$ FITC-labelled BS-IB4 and incubated at 4°C for 1 h. Cells were washed and then fixed as described above. FACS analysis was carried out using a fluorescence-activated cell sorter (FACScan, Beckton Dickinson). Histograms were drawn by using WinMDI software by Joseph Trotter.

TABLE 1 Inhibition of virus inactivation by Gal(α 1-3)Gal and pre-adsorption of human serum with Gal(α 1-3)Gal expressing cells

	Relative titre (% control)	
	lacZ(MLV-A)	lacZ(RD114)
a, Block of virus inactivation by sugar		
No sugar	0.15	<0.2
Gal α 1-3Gal (2 mg ml $^{-1}$)	34	<0.2
Gal α 1-3Gal (10 mg ml $^{-1}$)	75	11
Gal β 1-4Gal (2 mg ml $^{-1}$)	0.31	<0.2
Gal β 1-4Gal (10 mg ml $^{-1}$)	0.31	0.2
b, Adsorption of anti-cell antibodies		
Unadsorbed serum	<0.4	<0.06
Heat-inactivated serum	93	104
Serum adsorbed on NIH3T3 cells	14	29
+ heat-inactivated serum	<0.4	0.06
Serum adsorbed on Cf2ThS+L- cells	19	132
+ heat-inactivated serum	2.5	6.8
Serum adsorbed on HT-neo cells	ND	<0.06
Serum adsorbed on HG13 cells	ND	22
+ heat-inactivated serum	ND	0.1

a, Fresh human serum NHS-3 from an A+ healthy donor⁷ was used in all experiments. Helper-free lacZ(MLV-A) was produced by the murine GPE + Am12 packaging line and helper-positive lacZ(RD114) was produced by dog Cf2ThS+L- cells. Pseudotype stocks were harvested in serum-free opti-MEM as previously described^{7,28}. Gal(α 1-3)galactose (Gal α 1-3Gal) and Gal(β 1-4)galactose (Gal β 1-4Gal) (Dextra Laboratories) were dissolved in PBS. 20 μ l of virus harvest (660 infectious units (i.u.) for both lacZ(MLV-A) and lacZ(RD114)), 20 μ l of serum and 10 μ l of sugar solution or PBS were mixed on ice, then incubated for 1 h at 37°C. LacZ(MLV-A) and lacZ(RD114) were titrated on NIH3T3 and TE671 cells, respectively. The titres relative to control treatment with FCS plus corresponding sugar are shown. Similar results were obtained using a serum sample from another A+ donor, NHS-1⁷. b, A cDNA clone coding for porcine (α 1-3)GT, pSa13GT1, was inserted into the EcoRI site of the expression vector pcDNA3 (Invitrogen) containing a CMV promoter^{14,15}; certain porcine cells are sensitive to human serum lysis²⁹. HT1080 cells were transfected with this construct (pXsa13GT2.1.1) by calcium phosphate precipitation and G418 selection as described elsewhere³⁰. 6/24 colonies expressed Gal(α 1-3)Gal epitopes by FACS analysis after staining with FITC-BS-IB4 lectin (Sigma), one of which, HG13, was analysed further. HT-neo is a bulk population of G418 resistant cells made by transfection of empty pcDNA3 vector. Fresh human serum (0.2 ml) was incubated with 5×10^7 NIH3T3, Cf2ThS+L-, HG13 or HT-neo cells in the presence of 10 mM EDTA on ice for 1 h, separated from cells by filtration through 0.45 μ m filter, Spin-X (Costar) and then supplemented with 12 mM MgCl_2 and 12 mM CaCl_2 . More than 80% of complement activity was retained after cell adsorption, when measured by a haemolytic assay (total haemolytic complement assay kit, Binding Site Ltd). Helper positive lacZ(MLV-A) from murine NIH3T3 cells and lacZ(RD114) from Cf2ThS+L- cells were harvested in opti-MEM. Pre-adsorbed serum was mixed with an equal volume of either opti-MEM or heat-inactivated human serum as indicated, then virus (280 i.u. for lacZ(MLV-A); 3,000 i.u. for lacZ(RD114)) was added to the mixture (1:1) and incubated at 37°C for 1 h. Surviving lacZ pseudotype virus was titrated on TE671 cells. Titre relative to opti-MEM treated virus is shown. ND, not determined.

clone (HG13) showing bright staining with BS-IB4 (data not shown) could deplete antibodies responsible for virus inactivation from human serum (Table 1) and was sensitive to lysis by human serum (Fig. 2a). Furthermore, MLV-A and RD114 viruses produced from these cells expressed Gal(α 1-3)Gal epitopes on their envelopes (Fig. 2b). Such viruses were sensitive to inactivation by human serum that could be specifically inhibited with the Gal(α 1-3)galactose disaccharide (Fig. 2c). This inactivation of RD114 virus was blocked by pre-adsorption of anti-cell antibodies on dog cells or on the (α 1-3)GT-transfected human cells, but not on HT1080 cells without (α 1-3)GT (data not shown). Similar results were obtained following transfection of human TE671 cells and flow-sorting for BS-IB4 binding; both the cells and viruses derived from them were sensitive to human serum (data not shown).

Humans, apes and Old World monkeys lack a functional $\alpha(1-3)$ galactosyltransferase gene^{12,13} and do not express Gal($\alpha 1-3$)Gal terminal carbohydrate structures. These hosts are therefore not tolerized and can mount an antibody response against this epitope, probably stimulated by common bacteria of the gut^{17,18}. Antibodies recognizing Gal($\alpha 1-3$)Gal carbohydrate epitopes are abundant in human serum, produced by up to 1% of circulating B lymphocytes¹⁹. They may confer fortuitous immune protection to naive individuals by complement mediated lysis of certain bacteria¹⁷ and *Trypanosoma cruzi*²⁰. Gal($\alpha 1-3$)Gal expression by animal organs is thought to be a major barrier for xenotransplantation to human recipients, because 'natural' anti-Gal($\alpha 1-3$)Gal antibodies cause hyperacute rejection^{21,22}.

Our data demonstrate that two different C-type retroviruses produced by either mouse or dog cells were sensitive to human serum at least in part because of their expression of Gal($\alpha 1-3$)Gal epitopes. The previous conclusion that inactivation of retrovirus by human serum is antibody-independent relied on experiments using patient hypogammaglobulinaemic serum or mixture of

human antibody and guinea-pig or rabbit serum^{1,2}. Details of the experiments of hypogammaglobulinaemic serum were not presented and it is not clear whether the sera used might have contained low levels of anti-Gal($\alpha 1-3$)Gal antibodies. Further, Gal($\alpha 1-3$)Gal epitopes present in guinea-pig and rabbit serum might have inhibited the virus-killing function of human anti-Gal($\alpha 1-3$)Gal antibody. However, some retroviruses produced by human or primate cells have also been shown to be lysed or inactivated by fresh human serum^{2,7}, suggesting that at least one other mechanism must operate. Such an anti-Gal($\alpha 1-3$)Gal antibody independent mechanism might involve direct interaction of C1 molecules and virus envelope proteins^{3,6}.

Host-dependent expression of α -galactosyl epitopes on eastern equine encephalitis virus has been described²³ and lymphocytic choriomeningitis virus, Newcastle disease virus²⁴ and Sindbis virus²⁵ are differentially sensitive to human serum, dependent on the host producer cell. Some of these data might be explained by anti-Gal($\alpha 1-3$)Gal antibody triggering of complement²⁴. Sequence homology suggests that inter-species infection by

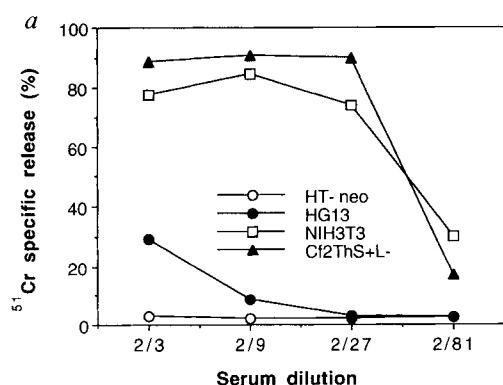
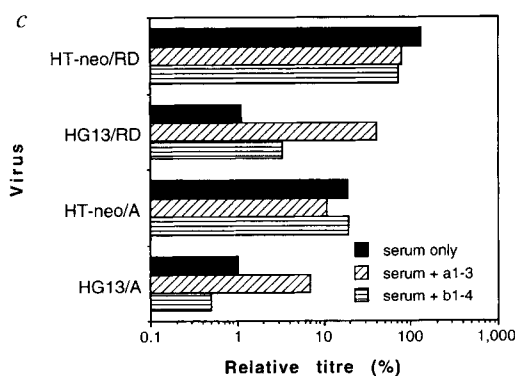
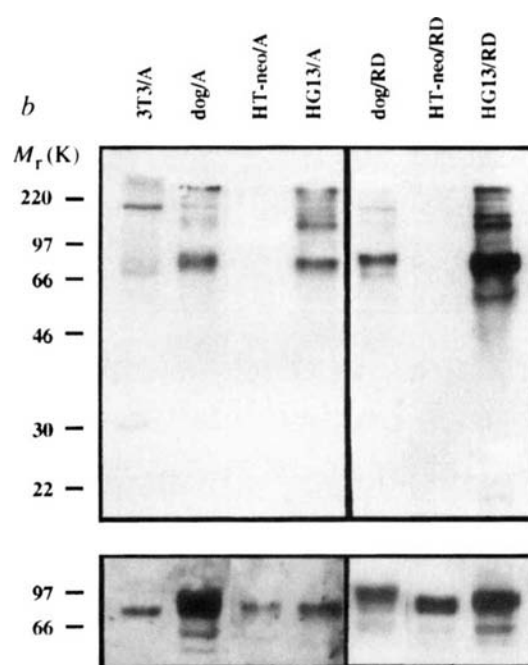


FIG. 2 $\alpha(1-3)$ Galactosyltransferase renders both cells and the retroviruses they produce sensitive to human serum. **a**, HT-neo and HG13 (Table 1), NIH3T3 and Cf2ThS+L- cells (2×10^6) were detached with EDTA and suspended in 200 μ l of [⁵¹Cr]sodium chromate (1 mCi ml^{-1} , Amersham) at 37°C for 1 h. After labelling, cells were washed and resuspended in DMEM with 10% FCS at 37°C for 30 min, then washed with serum-free opti-MEM and resuspended in serum-free opti-MEM at $2 \times 10^5 \text{ ml}^{-1}$. 50 μ l of cell suspension was mixed with 100 μ l of serum dilution (NHS-3) in a V-bottom microtitre well. Plates were incubated at 37°C for 3 h and the per cent specific ⁵¹Cr released into cell-free supernatant was determined by the following formula:

$$\left\{ \frac{\text{release with serum} - \text{release with serum-free medium}}{\text{release with 1\% NP40} - \text{release with serum-free medium}} \times 100 \right\}$$

$$\left\{ \frac{\text{release with 1\% NP40} - \text{release with serum-free medium}}{\text{release with 1\% NP40} - \text{release with serum-free medium}} \right\}$$

The mean specific ⁵¹Cr release of triplicates is shown. Standard errors were less than 3%. **b**, MFGnslacZ retroviral vector and replication competent virus MLV-A (A) or RD114 (RD) are introduced into NIH3T3 (3T3), Cf2ThS+L- (dog), HT-neo and HG13 cells as described previously^{7,8}. These cells persistently produced lacZ pseudotypes (lacZ titres on TE671 cells: 3T3/A, 2.5×10^5 ; dog/RD, 3.2×10^5 ; HT-neo/A, 6.2×10^5 ; HG13/A, 5.0×10^5 ; dog/RD, 4.8×10^5 ; HT-neo/RD, 8.5×10^5 ; HG13/RD, 7.2×10^5). Viral proteins were analysed by western blot as described previously²⁸. Briefly, viruses were harvested in DMEM supplemented with 2.5% FCS, pelleted by ultracentrifugation at 30,000 r.p.m. for 1 h and resuspended in loading buffer for SDS-polyacrylamide gel electrophoresis (10%). Upper panel; after blotting onto nitrocellulose, the filter was blocked for 1 h in 5% human serum albumin (HSA) (Sigma) in Tris-buffered saline (TBS) and then incubated overnight with 0.5 $\mu\text{g ml}^{-1}$ horseradish peroxidase (HRPO)-labelled BS-IB4 (Sigma) in TBS containing 5% HSA and 0.1% Tween-20. The filter was washed and developed using chemiluminescence (Amersham Life Science). Lower panels; the same filter was stripped, blocked in TBS containing 5% dry milk and reprobed with goat anti-Rauscher leukemia virus gp70-SU anti-serum (left panel) or anti-RD114 gp70-SU anti-serum (right panel) (Quality Biotech Inc, USA) at 1/1,000 and 1/5,000 respectively, followed by HRPO-labelled rabbit anti-goat Ig antibodies (DAKO, UK) at 1/2,000 and chemiluminescence. **c**, LacZ pseudotypes were harvested in serum-free opti-MEM for virus inactivation as described in Table 1a. Titres relative to opti-MEM treatment are shown for



lacZ(RD114) from HT-neo (HT-neo/RD) or HG13 (HG13/RD) and lacZ(MLV-A) from HT-neo (HT-neo/A) or HG13 (HG13/A) after treatment with fresh human serum (NHS-3) (serum only), serum plus 10 mg ml^{-1} Gal($\alpha 1-3$)galactose (serum + a1-3) and serum plus 10 mg ml^{-1} Gal($\beta 1-4$)galactose (serum + b1-4). Similar results were obtained using NHS-1.

C-type retroviruses has occurred frequently during mammalian evolution, but no human C-type retroviruses have to date been reported. The two C-type retroviruses, RD114 and MLV-A, can infect primary human lymphocytes in culture. Natural anti-Gal(α 1-3)Gal antibodies in human serum may therefore have provided one barrier, probably in combination with induced cellular immunity, to the transmission of a number of carbohydrate-carrying enveloped viruses to humans, apes and Old World monkeys from other mammals.

Our data also show that human cells expressing Gal(α 1-3)Gal carbohydrate modifications can be lysed by human serum, despite their expression of the complement inhibitory molecules DAF, CD46 and CD59 (determined by FACS analysis; data not shown). However, these transfected cells remain less sensitive to lysis by human serum than murine or canine cells (Fig. 2a). Therefore, lysis by human serum is probably controlled by a balance between the expression levels of complement inhibitory molecules and Gal(α 1-3)Gal carbohydrate epitopes. This suggests that the expression of Gal(α 1-3)Gal epitopes may need to be inhibited before animal organs (pig organs, for example) can be used efficiently in xenotransplantation to humans²⁶. Indeed, although human DAF and CD59 expression in transgenic pigs allowed prolonged heart graft survival in baboons, the organs were eventually rejected²⁷. □

Received 27 July; accepted 12 October 1995.

1. Welsh, R. M., Cooper, N. R., Jensen, F. C. & Oldstone, M. B. A. *Nature* **257** 612–614 (1975).

2. Welsh, R. M., Jensen, F. C., Cooper, N. R. & Oldstone, M. B. A. *Virology* **74**, 432–440 (1976).
3. Cooper, N. R., Jensen, F. C., Welsh, R. M. & Oldstone, M. B. A. *J. exp. Med.* **144**, 970–984 (1976).
4. Banapour, B., Sernatiner, J. & Levy, J. A. *Virology* **152**, 268–271 (1986).
5. Hoshino, H., Tanaka, H., Miwa, M. & Okada, H. *Nature* **310**, 324–325 (1984).
6. Bartholomew, R. M., Esser, A. F. & Muller-Eberhard, H. J. *J. exp. Med.* **147**, 844–853 (1978).
7. Takeuchi, Y. *et al. J. Virol.* **68**, 8001–8007 (1994).
8. Cosset, F. L., Takeuchi, Y., Battini, J. L., Weiss, R. A. & Collins, M. K. L. *J. Virol.* (in the press).
9. Barbacid, M., Bolognesi, D. & Aaronson, S. A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1617–1621 (1979).
10. Snyder, H. W. Jr & Fleissner, E. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1622–1626 (1980).
11. Löwer, J., *et al. Virology* **109**, 409–417 (1981).
12. Larsen, R. D., Rivera-Marrero, C. A., Ernst, L. K., Cummings, R. D. & Lowe, J. B. *J. biol. Chem.* **265**, 7055–7061 (1990).
13. Gallili, U. & Swanson, K. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7401–7404 (1991).
14. Gustafsson, K., Strahan, K. & Preece, A. *Immun. Rev.* **141**, 59–70 (1994).
15. Strahan, K. M. *et al. Immunogenetics* **41**, 101–105 (1995).
16. Sandrin, M. S. & McKenzie, I. F. *Immun. Rev.* **141**, 169–190 (1994).
17. Hamadeh, R. M. *et al. J. clin. Invest.* **89**, 1223–1235 (1992).
18. Gallili, U., Mandrell, R. E., Hamadeh, R. M., Shohet, S. B. & Griffiss, J. M. *Infect. Immun.* **56**, 1730–1737 (1988).
19. Gallili, U., Anaraki, F., Thall, A., Hill-Black, C. & Radic, M. *Blood* **82**, 2485–2493 (1993).
20. Almeida, I. C., Krautz, G. M., Kretzli, A. U. & Travassos, L. R. *J. clin. Lab. Anal.* **7**, 307–316 (1993).
21. Oriol, R., Ye, Y., Koren, E. & Cooper, D. K. *Transplantation* **56**, 1433–1442 (1993).
22. Sandrin, M. S., Vaughan, H. A., Dabkowski, P. L. & McKenzie, I. F. *Proc. natn. Acad. Sci. U.S.A.* **90**, 11391–11395 (1993).
23. Repik, P. M., Strizki, J. M. & Gallili, U. *J. gen. Virol.* **75**, 1177–1181 (1994).
24. Welsh, R. M. *J. Immun.* **118**, 348–354 (1977).
25. Hirsch, R. L., Griffin, D. E. & Winkelstein, J. A. *J. Immun.* **127**, 1740–1743 (1981).
26. Fabre, J. W. *Nature Med.* **1**, 403–404 (1995).
27. McCurry, K. R. *et al. Nature Med.* **1**, 423–427 (1995).
28. Cosset, F. L. *et al. J. Virol.* **69**, 6314–6322 (1995).
29. McKenzie, I. F. C. *et al. Lancet* **ii**, 386–387 (1968).
30. Battini, J. L., Heard, J. M. & Danos, O. *J. Virol.* **66**, 1468–1475 (1992).

ACKNOWLEDGEMENTS. We thank Claudia Flemming and Ian Tittle for expert technical assistance. This work was funded by the MRC and the CRC. BioTransplant Inc. in part supported K.M.S., A.F.P. and K.G. via a sponsored research agreement and kindly provided affinity purified human anti-Gal(α 1-3)Gal antibodies. K.G. was supported by the British Heart Foundation.

Hypoxia-mediated selection of cells with diminished apoptotic potential in solid tumours

Thomas G. Graeber*, Cynthia Osmanian*,
Tyler Jacks†‡, David E. Housman†,
Cameron J. Koch§, Scott W. Lowe†|| &
Amato J. Giaccia*¶

* Department of Radiation Oncology, Stanford University School of Medicine, Stanford, California 94305, USA

† Department of Biology, Center for Cancer Research, and

‡ Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

§ Radiation Oncology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

|| Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

Apoptosis is a genetically encoded programme of cell death that can be activated under physiological conditions^{1,2} and may be an important safeguard against tumour development^{3–6}. Regions of low oxygen (hypoxia) and necrosis are common features of solid tumours^{7,8}. Here we report that hypoxia induces apoptosis in oncogenically transformed cells and that further genetic alterations, such as loss of the *p53* tumour-suppressor gene or over-expression of the apoptosis-inhibitor protein Bcl-2, substantially reduce hypoxia-induced cell death. Hypoxia also selects for cells with defects in apoptosis, because small numbers of transformed cells lacking *p53* overtake similar cells expressing wild-type *p53* when treated with hypoxia. Furthermore, highly apoptotic regions strongly correlate with hypoxic regions in transplanted tumours expressing wild-type *p53*, whereas little apoptosis occurs in hypoxic regions of *p53*-deficient tumours. We propose that

TABLE 1 *p53*-modulated hypoxic induction of apoptosis *in vivo*

	Number of apoptotic cells per 0.01 mm ²		Fold increase
	Aerobic regions	Hypoxic regions	
<i>p53</i> /			
<i>p53</i> ^{+/+}	4.1 ± 1.8	7.9 ± 4.4	1.9
Ratio	3.7 ± 1.7	26.8 ± 9.4	7.2
	0.9	3.4 (6.1*)	
+/+ to -/-			

Quantification of apoptosis was performed on the TUNEL and EF5 stained sections described in Fig. 3. Fields containing both aerobic (EF5-negative) and hypoxic (EF5-positive) regions were randomly chosen while blinded to the apoptosis staining results (see, for example, Fig. 3d) and then photographed. The number of apoptotic cells per unit area was then calculated for each distinct region. Values stated are means ± s.d. (*n* = 16, 4 tumours each genotype).

* With aerobic values subtracted.

hypoxia provides a physiological selective pressure in tumours for the expansion of variants that have lost their apoptotic potential, and in particular for cells acquiring *p53* mutations.

Solid tumours that possess low-oxygen regions have a poorer prognosis than well oxygenated tumours, independent of treatment⁹. To investigate how oncogenically transformed cells respond to hypoxia, Rat1 fibroblasts constitutively expressing a c-Myc-oestrogen receptor chimera protein (MycER), which is activated by 4-hydroxytamoxifen (4hT)¹⁰, were cultured either aerobically or anaerobically in the presence of 4hT and 10% serum for 48 h. Hypoxia induced substantial apoptosis (85 ± 2%; Fig. 1b) as compared to aerobic culture (10 ± 2; Fig. 1a). Bcl-2, which prevents c-Myc-mediated apoptosis following serum depletion³, also blocked c-Myc-mediated apoptosis induced by hypoxia (Fig. 1c–d,i). These data indicate that oncogenic changes can modulate the susceptibility of cells to hypoxia-induced apoptosis.

p53 is an integral part of DNA damage-induced apoptotic pathways¹¹. Because cells exposed to hypoxia accumulate wild-type

¶ To whom correspondence should be addressed.

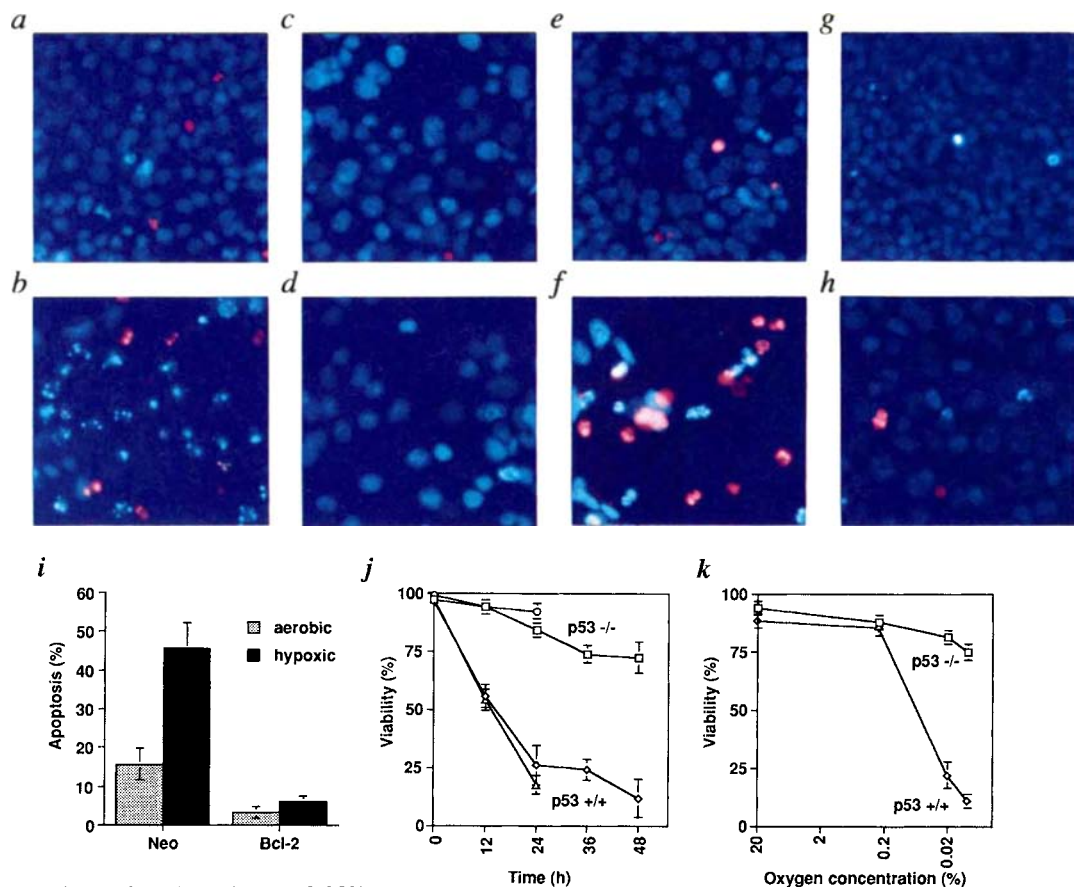
p53 protein¹², we investigated the role of p53 in hypoxia-induced apoptosis using embryonic fibroblasts derived from wild-type and p53-deficient transgenic mice (Fig. 1*e–h, j*). Transformation of p53-expressing mouse embryonic fibroblasts (MEFs) by the adenovirus early region 1A (E1A) and activated Ha-ras oncogenes dramatically enhanced their susceptibility to hypoxia-induced apoptosis. Untransformed MEFs remained over 95% viable, whereas their transformed derivatives were only 11% viable after 48 h of hypoxia. p53 protein accumulated under hypoxia (not shown) and was required for efficient induction of apoptosis because oncogenically transformed cells lacking p53 were resistant to hypoxia-induced killing (72% viability after 48 h; Fig. 1*j*). Nonetheless, the p53^{-/-} cells were also dying through apoptosis, indicating the existence of a less sensitive p53-independent apoptotic pathway³. The influence of p53 on apoptotic potential was not apparent until oxygen concentrations were reduced below 0.2% (Fig. 1*k*), corresponding to concentrations commonly found in solid tumours^{8,9}. Thus, oncogenic transformation predisposes cells to hypoxia-induced killing through an apoptotic pathway modulated by p53.

The difference in viability of oncogenically transformed cells that differ in p53 status suggests that p53-deficient cells would

have a survival advantage over wild-type p53 cells in low-oxygen conditions. To test this hypothesis, we mixed transformed p53^{-/-} (E1A, Ha-ras) cells stably expressing the lacZ gene product (lacZ⁺) (Fig. 2*c*) with p53^{+/+} (E1A, Ha-ras) lacZ⁻ cells at a 1 to 1,000 ratio and treated them with multiple rounds of hypoxia and aerobic recovery. The percentage of p53^{-/-} cells increased by approximately 2.4-fold following each treatment, and after 7 treatments the p53^{-/-} (lacZ⁺) cells had overtaken the p53^{+/+} cells (Fig. 2*d–f*). Control cells (Fig. 2*a*) grown aerobically in parallel had no detectable increase in their percentage of p53^{-/-} cells (Fig. 2*b*). These experiments demonstrate that hypoxia can select for apoptosis-resistant cells.

To investigate the relationship between oxygen concentration, p53 status and apoptosis *in vivo*, we determined whether apoptotic regions correlated with hypoxic regions in tumours derived from the E1A- and Ha-ras-transformed p53^{+/+} and p53^{-/-} cells¹³. Apoptosis was assessed using the terminal deoxynucleotidyl transferase-mediated deoxyuridine nick end-labelling (TUNEL) assay. In TUNEL-stained sections, apoptotic regions were more prevalent in p53^{+/+} tumours than in p53^{-/-} tumours (see below), and apoptosis was most pronounced in regions distal to adjacent blood vessels (Fig. 3*a, b*).

FIG. 1 Apoptotic response of oncogenically transformed cells to hypoxia. Following aerobic (*a, c, e, g*) or anaerobic (*b, d, f, h*) culture, nuclear morphology was visualized by staining with bisbenzamide (blue) and apoptotic cells were identified by their characteristic nuclear condensation and fragmentation¹. Loss of membrane integrity was assessed by permeability to propidium iodide (PI; pink). Cell lines shown are MycER/Rat1 (*a, b*), MycER/Rat1 + Bcl-2 (*c, d*), p53^{+/+} (E1A, Ha-ras) (*e, f*) and p53^{-/-} (E1A, Ha-ras) (*g, h*) at $\times 270$ magnification. Rat1 (p53^{+/+}) cells transformed by E1A and Ha-ras also underwent apoptosis when cultured under hypoxic conditions. *i*, Quantification of the percentage of apoptotic cells in control transfectants (Neo) or Bcl-2 transfectants of MycER/Rat1 cells cultured aerobically or under hypoxic conditions. Results shown in *i* are the means $\pm$ s.e.m. of two independently derived clones. *j, k*, Loss of viability due to induced apoptosis in p53^{+/+}



(E1A, Ha-ras) and p53^{-/-} (E1A, Ha-ras) cells following culture at 0.02% oxygen (*j*), or following 24 h at various oxygen concentrations (*k*). Data shown in (*j, k*) are the means $\pm$ s.e.m. of two independent experiments. Each curve in (*j*) represents data from an independently derived clone. Untransformed p53^{-/-} MEFs were 95% viable after 48 h of hypoxia (0.02% oxygen).

METHODS. MycER/Rat1 fibroblasts were generated by transfection of Rat1 fibroblasts with the plasmid pBpuro c-mycER which encodes for a MycER chimera protein that is activated by 4hT¹⁰. MycER/Rat1 cells expressing Bcl-2 protein (verified by immunoblot analysis) were generated by subsequent transfections with the plasmid pSFFV Bcl-2n²⁷. To activate Myc, 100 nM 4hT was added to the media of control and hypoxia-treated cells at the start of each experiment. MycER/Rat1 cells treated with hypoxia in the

absence of 4hT showed similar low levels of apoptosis as hypoxia-treated parental Rat1 cells. Transformed p53^{+/+} MEFs which express E1A and Ha-ras were generated as described²⁸. Although p53 promotes apoptosis induced by hypoxia, p53 is not required for hypoxia-induced cell-cycle block¹². Hypoxic incubations were performed in anaerobic incubators as described¹², with an oxygen concentration of 0.02% for 48 h unless otherwise noted. Before microscopy, cells were incubated with 5 μ g ml⁻¹ each of bisbenzamide (Hoechst No. 33342) and PI (*a–i*). Apoptosis/viability ratios were determined by scoring low-magnification photographs of randomly selected fields for cells with condensed and fragmented nuclear morphology (*i*), or by trypan-blue dye exclusion (*j, k*). All cells were cultured in media containing 10% fetal calf serum.

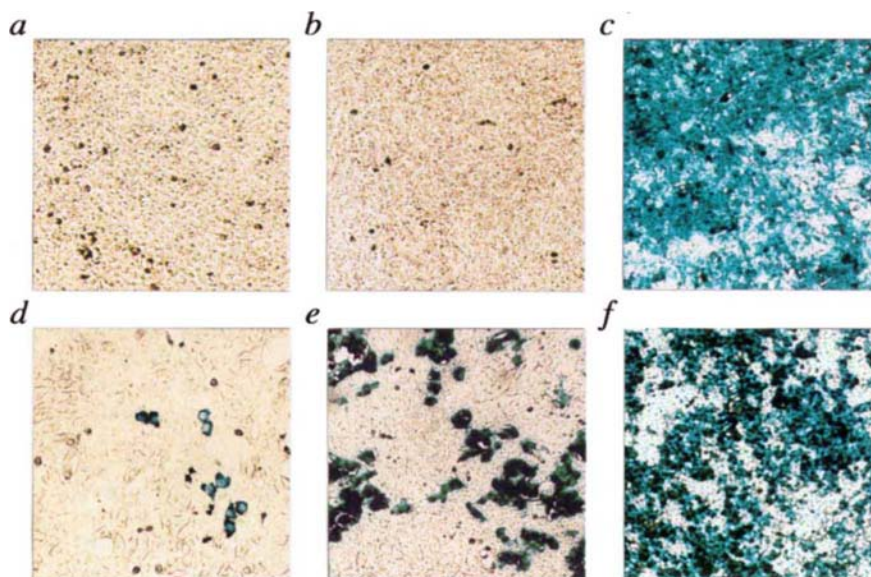


FIG. 2 Selection for $p53^{-/-}$ cells following hypoxia treatment *in vitro*. $p53^{-/-}$ (E1A, Ha-ras) cells stably expressing the *lacZ* gene product (*lacZ*⁺; blue) were mixed with $p53^{+/+}$ (E1A, Ha-ras) cells (*lacZ*⁻; clear) at a ratio of 1 to 1,000 and treated with multiple rounds of hypoxia and aerobic recovery. Representative photographs of β -galactoside-stained cells are shown at $\times 135$ magnification: a, mixed cells at start of experiment; b, mixed cells cultured aerobically in parallel for the duration of the experiment; c, 100% $p53^{-/-}$ (*lacZ*⁺) cells, and mixed cells after three (d), six (e) and seven (f)

rounds of hypoxia treatment.

METHODS. $p53^{-/-}$ (*lacZ*⁻) populations were generated by viral infection of $p53^{+/+}$ (E1A, Ha-ras) cells using the MFG-*lacZ* retrovirus²⁹, with subsequent enrichment for *lacZ*⁻ cells by fluorescence-activated cell sorting³⁰. Treated mixed cells were subjected to multiple rounds of hypoxia (2–3 days) followed by recovery (3–5 days) with sub-culturing when confluent. *LacZ*⁺ cells were detected by staining for β -galactosidase activity³⁰.

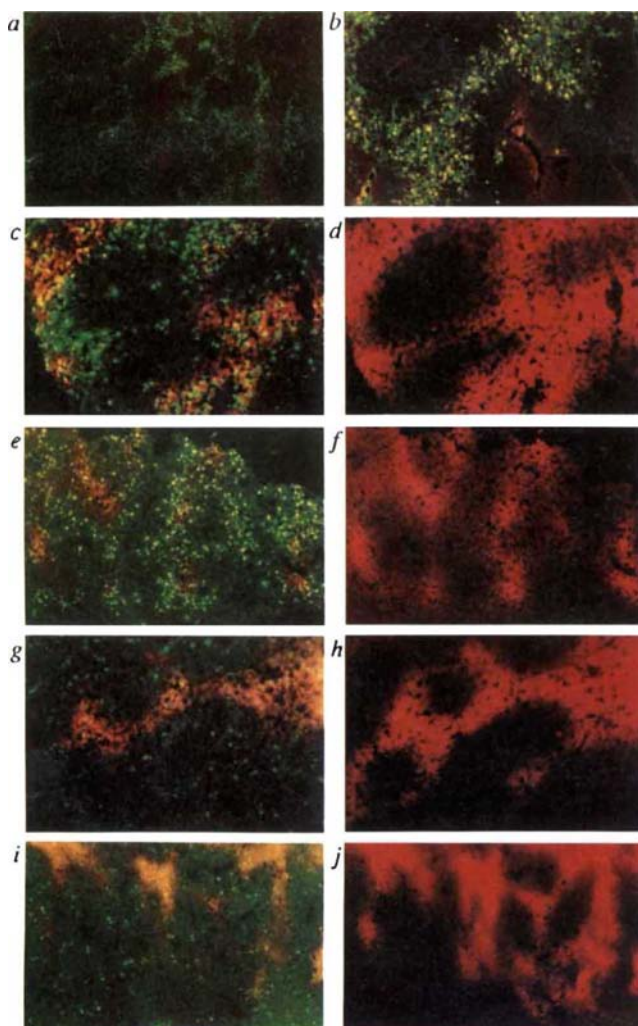


FIG. 3 Relationship between hypoxic and apoptotic regions in tumours differing in $p53$ status. a, b, TUNEL staining (green) of a $p53^{-/-}$ paraffin tumour section showing apoptotic regions distal to blood vessels (counter-stained with propidium iodide (red); a, magnification $\times 25$; b, $\times 135$). c–j, TUNEL staining (green) and EF5 staining (orange/red) of frozen tumour sections comparing the extent of apoptosis in similarly stained EF5-positive hypoxic regions of $p53^{-/-}$ tumours (c–f) and $p53^{+/+}$ tumours (g–j). Photographs c, e, g, i simultaneously show both TUNEL staining and the brightest EF5 staining using a dual-band pass fluorescence filter set, and photographs d, f, h, j show all EF5 staining of the same regions using a single-band pass filter set (c, d, g, h, magnification $\times 135$; e, f, i, j, $\times 85$). Representative photographs of multiple tumours are shown.

METHODS. $p53^{+/+}$ (E1A, Ha-ras) and $p53^{-/-}$ (E1A, Ha-ras) cells were transplanted subcutaneously into the backs of immunocompromised mice as described¹³ (a, b, athymic nude mice; c–j, severe combined immunodeficient (SCID) mice). After tumours reached a volume of 0.2 to 1.2 cm³, the tumours were either excised and fixed for paraffin section processing (a, b; 5 μ m thick), or the mice were given an intraperitoneal injection of EF5 (2(2-nitro-1H-imidazol-1-yl)-N-(2,2,3,3,3-pentafluoropropyl)-acetamide) 1 h before tumour excision and freezing for frozen sectioning (c–j; 15 μ m thick) as described¹⁴. In the absence of oxygen, EF5 is metabolically activated in viable cells and forms adducts with cellular macromolecules which can be detected with a monoclonal antibody¹⁴. TUNEL staining (fluorescein, green) was done according to the manufacturer's instructions (Oncor). Where applicable, EF5 staining (Cy3 fluorochrome, red) was done subsequently on the same section as described¹⁴. Hypoxic regions were present in all tumour sizes. Frozen sections of control tumours that were not exposed to EF5 had similar apoptotic patterning as tumours shown in c–j. 1-h exposures of aerobic or hypoxic cells to EF5 did not induce apoptosis *in vitro*.

To determine directly whether p53 influences the magnitude of apoptosis in tumour regions of similar oxygen deficiency, the specific hypoxia marker EF5¹⁴ was injected 1 h before excision of p53^{+/+} and p53^{-/-} tumours. Frozen sections from these tumours were then stained by the TUNEL assay and with anti-EF5 antibody. Although no difference in hypoxic fraction or patterning was observed between tumours derived from the two p53 genotypes, p53^{+/+} tumours showed a more distinct spatial correlation between apoptotic and hypoxic regions compared to p53^{-/-} tumours (Fig. 3c–j). The frequency of apoptosis was 3.4-fold greater in hypoxic regions of p53^{+/+} tumours than in hypoxic regions of p53^{-/-} tumours, whereas no difference was seen in aerobic regions from the same tumours. Hypoxic regions in p53^{+/+} tumours displayed 7.2 times more apoptosis than aerobic regions (Table 1). *In vivo*, other tumour-microenvironmental factors determined by vascularization and angiogenesis could act to induce apoptosis synergistically, although *in vitro* data indicate that hypoxia is sufficient (Fig. 1).

The data presented here have important implications for multi-step carcinogenesis. Oncogenic changes occurring early in solid tumour development promote deregulated proliferation and can also increase cellular susceptibility to apoptosis^{15–17}. We have shown that oncogenic transformation increases the sensitivity of cells to hypoxia and that hypoxia correlates spatially with apoptosis *in vivo*. The same genetic events that suppress hypoxia-induced apoptosis in oncogenically transformed cells also temporally coincide with reduced apoptosis in later stages of tumour development^{5,18–21}. These observations suggest that hypoxia acts as a physiological selective agent against apoptosis-competent cells in tumours, thus promoting the clonal expansion of cells that acquire mutations in their apoptotic programmes.

p53 mutation and Bcl-2 overexpression also suppress apoptosis induced by radiation and chemotherapy^{22–24}. Hence, hypoxia-mediated selection of cells with diminished apoptotic potential could also explain the resistance of many solid tumours to cancer therapy²⁴. Although apoptosis induced by ultraviolet radiation has been implicated in the clonal selection of p53 mutations in skin cancer²⁵, most solid tumours are not exposed to ultraviolet radiation during their development. In contrast, hypoxia is a physiological stress commonly found in developing solid tumours of diverse origins⁸, and thus our results offer an explanation for why p53 is one the most commonly mutated genes in human cancer²⁶. □

Received 28 July; accepted 25 October 1995.

1. Bellamy, C. O. C., Malcolmson, R. D. G., Harrison, D. J. & Wyllie, A. H. *Semin. Cancer Biol.* **6**, 3–16 (1995).
2. Thompson, C. B. *Science* **267**, 1456–1462 (1995).
3. Harrington, E. A., Fanidi, A. & Evan, G. I. *Curr. Opin. Genet. Dev.* **4**, 120–129 (1994).
4. Wyllie, A. H. *Curr. Opin. Genet. Dev.* **5**, 97–104 (1995).
5. Symonds, H. et al. *Cell* **78**, 703–711 (1994).
6. Deng, C., Zhang, P., Harper, J. W., Elledge, S. J. & Leder, P. *Cell* **82**, 675–684 (1995).
7. Thomlinson, R. H. & Gray, L. H. *Br. J. Cancer* **9**, 539–549 (1955).
8. Vaupel, P. W. & Hockel, M. in *Tumor Oxygenation* (eds Vaupel, P. W., Kelleher, D. K. & Gunderoth, M.) 219–232 (Gustav Fischer, Stuttgart, 1995).
9. Hockel, M., Vorndran, B., Schlenger, K., Baussmann, E. & Knapstein, P. G. *Gynecol. Oncol.* **51**, 141–149 (1993).
10. Littlewood, T. D., Hancock, D. C., Danielian, P. S., Parker, M. G. & Evan, G. I. *Nucleic Acids Res.* **23**, 1686–1690 (1995).
11. Lane, D. P., Lu, X., Hupp, T. & Hall, P. A. *Phil. Trans. R. Soc. B* **345**, 277–280 (1994).
12. Graeber, T. G. et al. *Molec. cell. Biol.* **14**, 6264–6277 (1994).
13. Lowe, S. W. et al. *Science* **266**, 807–810 (1994).
14. Lord, E. M., Harwell, L. & Koch, C. J. *Cancer Res.* **53**, 5721–5726 (1993).
15. Bishop, J. M. *Science* **235**, 305–311 (1987).
16. Harrington, E. A., Bennett, M. R., Fanidi, A. & Evan, G. I. *EMBO J.* **13**, 3286–3295 (1994).
17. Lowe, S. W., Ruley, H. E., Jacks, T. & Housman, D. E. *Cell* **74**, 957–967 (1993).
18. Bardeesy, N., Beckwith, J. B. & Pelletier, J. *Cancer Res.* **55**, 215–219 (1995).
19. Baker, S. J. et al. *Cancer Res.* **50**, 7717–7722 (1990).
20. Sinicrope, F. A. et al. *Cancer Res.* **55**, 237–241 (1995).
21. Bedi, A. et al. *Cancer Res.* **55**, 1811–1816 (1995).
22. Lotem, J. & Sachs, L. *Cell Growth Differ.* **4**, 41–47 (1993).
23. Lowe, S. W. *Curr. Opin. Oncol.* (in the press).
24. Hickman, J. A., Potten, C. S., Merritt, A. J. & Fisher, T. C. *Phil. Trans. R. Soc. B* **345**, 319–325 (1994).
25. Ziegler, A. et al. *Nature* **372**, 773–776 (1994).
26. Greenblatt, M. S., Bennett, W. P., Hollstein, M. & Harris, C. C. *Cancer Res.* **54**, 4855–4878 (1994).
27. Hockenbery, D., Nunez, G., Millman, C., Schreiber, R. D. & Korsmeyer, S. J. *Nature* **348**, 334–336 (1990).
28. Lowe, S. W., Jacks, T., Housman, D. E. & Ruley, H. E. *Proc. natn. Acad. Sci. U.S.A.* **91**, 2026–2030 (1994).

29. Dranoff, G. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 3539–3543 (1993).

30. Pear, W., Nolan, G. P., Scott, M. L. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **90**, 8392–8396 (1993).

ACKNOWLEDGEMENTS. We thank B. Sim, D. Rapacchietta, D. Menke and M. Tsai for technical assistance, M. Brown for helpful discussions, G. Evan for the pBpuro c-mycER plasmid, M. McMahon (DNAX Research Institute) for 4hT, D. Hockenbery for the pSFFV Bcl-2nl plasmid and anti-Bcl-2 antibody and R. Mulligan for the MFG-lacZ retrovirus. T.G.G. is a Fannie and John Hertz Foundation predoctoral fellow and S.W.L. is an Anna Fuller postdoctoral fellow. This work was supported by grants from the NCI and ACS to A.J.G. and S.W.L.

Regulation of cell adhesion and anchorage-dependent growth by a new β_1 -integrin-linked protein kinase

Gregory E. Hannigan^{*†‡}, Chungyee Leung-Hagesteijn^{*}, Linda Fitz-Gibbon^{*}, Marc G. Coppelino^{*†}, Galina Radeva^{*†}, Jorge Filmus^{*†}, John C. Bell[§] & Shoukat Dedhar^{*†‡}

^{*} Cancer Biology Research Program, Sunnybrook Health Science Centre, and [†] Department of Medical Biophysics, University of Toronto, 2075 Bayview Avenue, Toronto M4N 3M5, Ontario, Canada

[§] Department of Biochemistry, University of Ottawa, Ottawa K1N 6N5, Ontario, Canada

THE interaction of cells with the extracellular matrix regulates cell shape, motility, growth, survival, differentiation and gene expression, through integrin-mediated signal transduction^{1–3}. We used a two-hybrid screen to isolate genes encoding proteins that interact with the β_1 -integrin cytoplasmic domain. The most frequently isolated complementary DNA encoded a new, 59K serine/threonine protein kinase, containing four ankyrin-like repeats. We report here that this integrin-linked kinase (ILK) phosphorylated a β_1 -integrin cytoplasmic domain peptide *in vitro* and coimmunoprecipitated with β_1 in lysates of mammalian cells. Endogenous ILK kinase activity was reduced in response to fibronectin. Overexpression of p59^{ILK} disrupted epithelial cell architecture and inhibited adhesion to integrin substrates, while inducing anchorage-independent growth. We propose that ILK is a receptor-proximal protein kinase regulating integrin-mediated signal transduction.

A partial cDNA, BIT-9, was isolated in a two-hybrid screen⁴ using a bait plasmid expressing the cytoplasmic domain of the β_1 integrin subunit. The BIT-9 insert was used to isolate clones from a human placental cDNA library. A 1.8 kilobase (kb) clone, Plac5, was found to contain a high degree of similarity to cDNAs encoding protein kinases (Fig. 1a–c), and recognized a widely expressed transcript of 1.8 kb in northern blots (Fig. 1d). Deduced amino-acid residues 186–451 from Plac5 comprise a domain which is highly homologous with the catalytic domains of a large number of protein tyrosine and serine/threonine kinases (Fig. 1b). Residues 33–164 comprise four repeats of a motif originally identified in erythrocyte ankyrin⁵ (Fig. 1c), probably defining a domain involved in mediating additional protein–protein interactions^{6,7}. Affinity-purified anti-ILK antibodies (see methods, Fig. 3) were used in western blot analyses of mammalian cell extracts, and detected a conserved protein of apparent M_r of 59K (p59^{ILK}, Fig. 1e).

For analysis of kinase activity *in vitro*, a bacterially expressed fusion protein, GST-ILK¹³², was band-purified after sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE), and incubated with [γ -³²P]ATP in the presence or absence of the exogenous substrate myelin basic protein (Fig. 2). GST-ILK¹³²

[‡]Present address: Section in Molecular Pathology, The Hospital for Sick Children, 555 University Avenue, Toronto M5G 1X8, Canada

| To whom correspondence should be addressed.

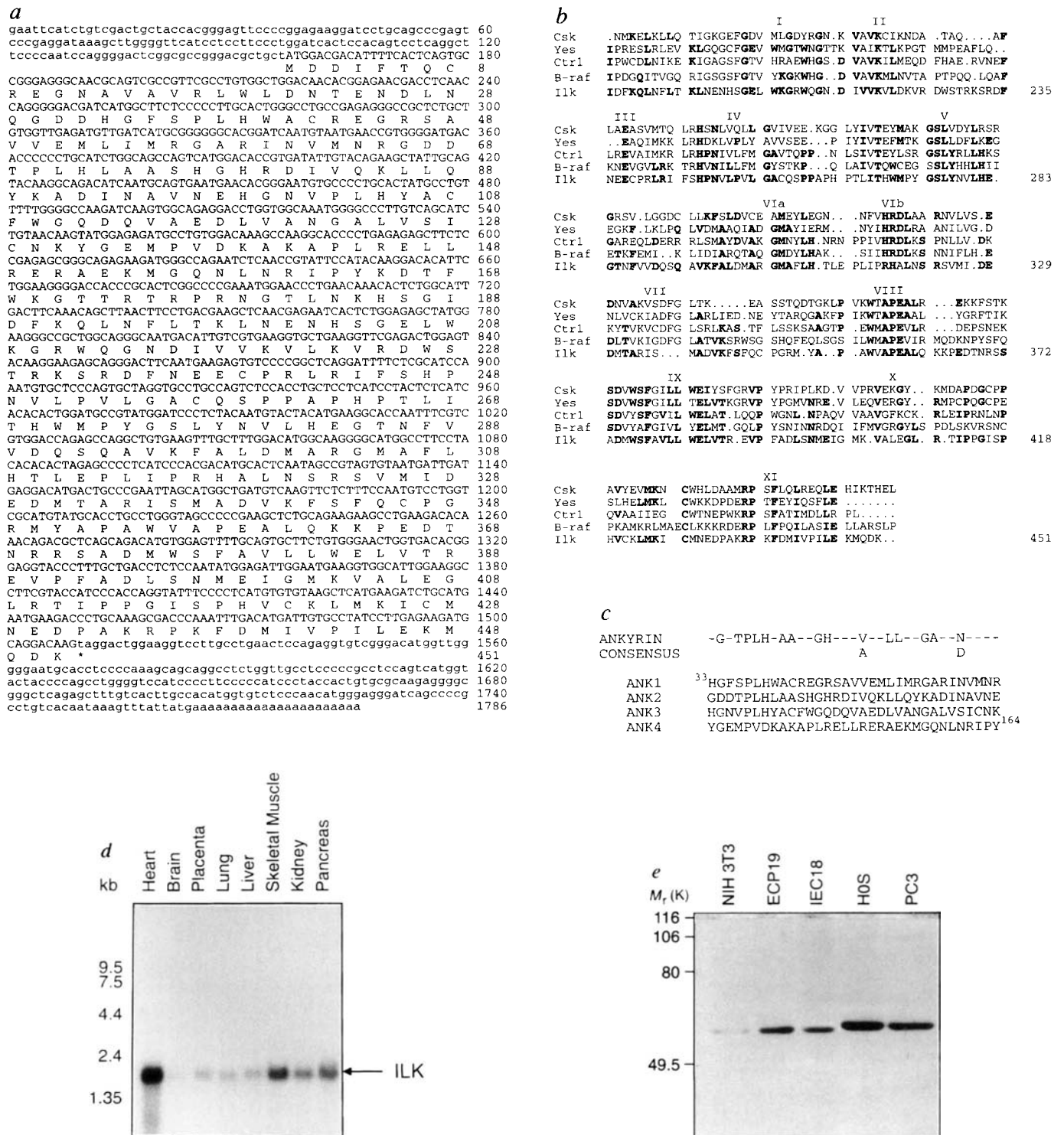


FIG. 1 Yeast two-hybrid cloning, characterization, and expression of ILK. **a**, The full-length ILK cDNA, Plac5, was isolated from a human placental library using the BIT-9 insert. Plac5 contains a 1,509bp open reading frame, with a presumptive initiator Met¹⁷ at nucleotide (nt) 157, and an AAUAAA signal 11bp upstream of the polyadenylation site. *in vitro* transcription and translation of Plac5 in rabbit reticulocyte lysates yielded a protein of apparent *M_r* of 59K (not shown). **b**, A search¹⁸ of the PIR protein database indicated homology with protein kinase subdomains I to XI, as identified in ref. 19. We note sequence variations in the ILK subdomains I, VIb and VII, relative to catalytic domains of known protein kinases. Subdomain I (residues 199–213), does not have the typical GXGXXG motif, although this region in ILK is Gly-rich. In subdomain VIb, Asp 328 of ILK may compensate for the lack of the otherwise conserved Asp 319. In subdomain VII, the DFG triplet is absent in ILK. The integrin binding site

maps to amino-acid residues 293–451 (BIT-9). The ILK kinase domain is most highly related to the CTR1 kinase of *Arabidopsis thaliana* (30% identity, $P < 10^{-13}$). The CTR1, B-raf, Yes and Csk kinase domains are aligned with Plac5. **c**, Amino-acid residues 33–164 comprise four contiguous ankyrin repeats, as defined in ref. 5. **d**, BIT-9 was used to probe a blot of poly(A)⁺ selected RNA (MTN I, Clontech) from various human tissues, **e**, Whole-cell lysates of mouse, rat and human cell lines (10 µg per lane) were analysed by western blotting with the affinity-purified 92-2 antibody (see Fig. 3, Methods). The ILK sequence data are available from GenBank under accession number U40282.

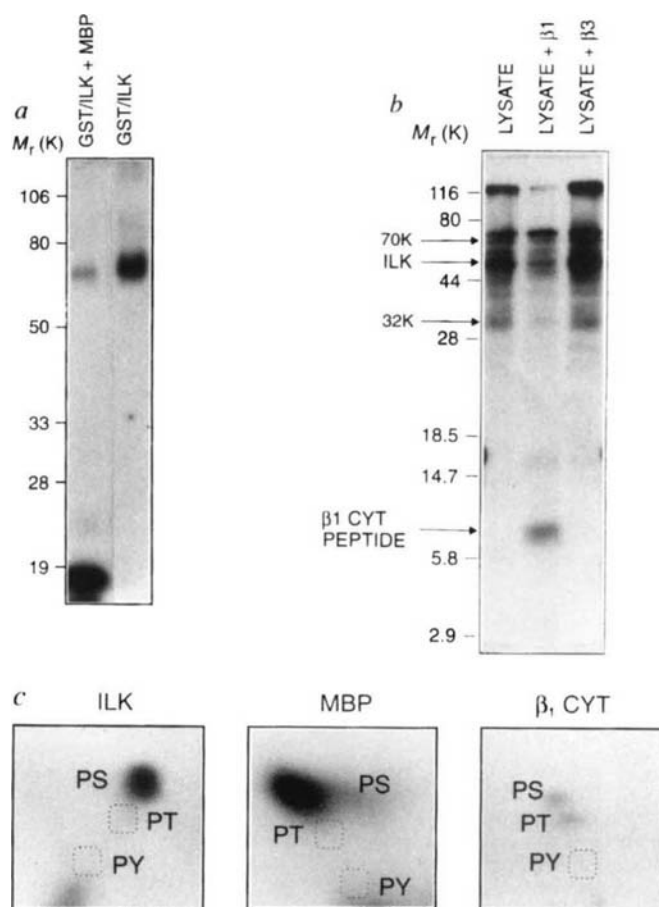
METHODS. To construct integrin 'bait' plasmids²⁰, sequences encoding amino-acid residues 738–798 of the β_1 , and residues 1,022–1,049 of the α_5 integrin subunits were amplified from full-length cDNAs²¹. PCR products were directionally cloned into pEG202, creating the LexA fusion bait

FIG. 2 *In vitro* and immune-complex kinase assays. *a*, *In vitro* kinase reactions containing 2 μ g of gel-purified GST-ILK¹³², with and without 5 μ g of myelin basic protein (MBP, Upstate Biotechnologies), were analysed by 10% SDS-PAGE. *b*, Immune complexes were generated from PC3 whole-cell lysates, using affinity-purified 91-3 antibody. Complexes were assayed for kinase activity, with and without addition of 5 μ g per reaction of synthetic peptides, representing β_1 or β_3 integrin cytoplasmic domains²⁴, or MBP (not shown). Products were analysed by 15% SDS-PAGE (M_r markers at left), and migration of peptides confirmed by Coomassie blue staining. *c*, ³²P-labelled products from the anti-ILK immune complex kinase reactions shown in *b*, were isolated and analysed for phosphoamino-acid content. Anti-FAK^{8,9} immune complex kinase assays demonstrated phosphotyrosine on MBP (not shown).

METHODS. Protein kinase assays were performed in 50- μ l kinase reaction buffer (50 mM HEPES pH 7.0, 10 mM MnCl₂, 10 mM MgCl₂, 2 mM NaF, 1 mM Na₃VO₄), containing 10 μ M [γ -³²P]ATP. Reactions were incubated at 30 °C for 20 min, and stopped by the addition of SDS-PAGE sample buffer. For assay of recombinant ILK activity, GST-ILK¹³² was adsorbed from bacterial lysates onto glutathione-agarose beads, or GST-ILK¹³² was band-purified from 10% SDS-PAGE gels. For immune complex kinase assays, affinity-purified 91-3 anti-ILK antibody (Fig. 3, Methods) was used to generate immunoprecipitates from Nonidet-P40 (NP-40) lysates (150 mM NaCl, 1% (v/v) NP-40, 0.5% (w/v) sodium deoxycholate, 50 mM HEPES pH 7.5, 1 μ g ml⁻¹ each leupeptin and aprotinin, 50 μ g ml⁻¹ phenyl-methylsulphonyl fluoride) of PC3 cells. Kinase reaction products were resolved on 10–15% SDS-PAGE gels, transferred to polyvinylidene fluoride, and phosphoamino-acid analysis performed according to a published protocol²⁵.

autophosphorylated and labelled myelin basic protein (MBP) efficiently in these assays (Fig. 2*a*). Anti-GST-ILK¹³² (antibody 91-3) immunoprecipitates of PC3 cell lysates were incubated with [γ -³²P]ATP, similar to experiments performed with purified recombinant GST-ILK¹³². ILK immune complexes labelled a protein of apparent M_r of 59K (Fig. 2*b*), corresponding to p59^{ILK}, as well as cellular proteins of apparent M_r 32K and 70K, which may be endogenous ILK substrates (Fig. 2*b*). We also see cellular phosphoproteins (serine/threonine) of about 32K and 70K, in β_1 -integrin-specific immune complex kinase assays (not shown).

In ILK immune complex kinase assays, a synthetic peptide representing the β_1 cytoplasmic domain was phosphorylated, whereas a similar peptide representing the β_3 cytoplasmic domain was not detectably labelled by p59^{ILK}. The β_1 peptide selectively inhibited autophosphorylation of ILK in these reactions (Fig. 2*b*), further indicating a differential interaction



of the peptides with ILK. The results demonstrating phosphorylation of synthetic β -peptides by endogenous ILK are identical to those seen with recombinant GST-ILK³² (not shown), and indicate the potential substrate preference of ILK for the β_1 cytoplasmic tail. This does not, however, necessarily rule out an interaction between ILK and the β_3 -integrin cytoplasmic domain. Phosphoamino-acid analyses of labelled p59^{ILK} and MBP from the immune complex kinase assays detected only phosphoserine in both substrates (Fig. 2*c*), as was the case for phosphorylation of these substrates by GST-ILK¹³² (not shown). The β_1 peptide was labelled on serine and threonine residues, with roughly equal stoichiometry (Fig. 2). As a control, anti-FAK^{8,9} immune complexes from the same lysates were analysed for phosphorylation of MBP, and phosphotyrosine was readily detected (not shown).

Immunofluorescence experiments indicated that ILK and β_1 integrin colocalize in focal plaques (not shown). To test further for this association in intact mammalian cells, we performed co-immunoprecipitation assays in lysates of PC3 cells, in which integrin expression has been well characterized¹⁰. PC3 cell lysates were immunoprecipitated with specific anti-integrin antibodies¹¹, and immune complexes analysed by western blotting with the anti-ILK antibody, 92-2. The specificities of the anti-ILK antibodies were tested by immunoprecipitation and western blotting (Fig. 3*a, b*). We detected p59^{ILK} in immune complexes obtained with anti-fibronectin receptor (FNR, $\alpha_5/\alpha_3\beta_1$ integrin), and anti-vitronectin receptor (VNR, $\alpha_v\beta_3/\beta_5$ integrin) antibodies, but not in those obtained with non-immune serum (Fig. 3*c*). Three anti- β_1 monoclonal antibodies also coprecipitated p59^{ILK} from PC3 lysates, confirming the β -integrin specificity of p59^{ILK} interaction (Fig. 3*d*). The detection of p59^{ILK} in anti-VNR immune complexes suggests that ILK may also interact with the β_3 and/or β_5 integrin subunit(s).

We next tested for fibronectin-dependent regulation of ILK kinase activity. Plating of rat intestinal epithelial cells, IEC-18

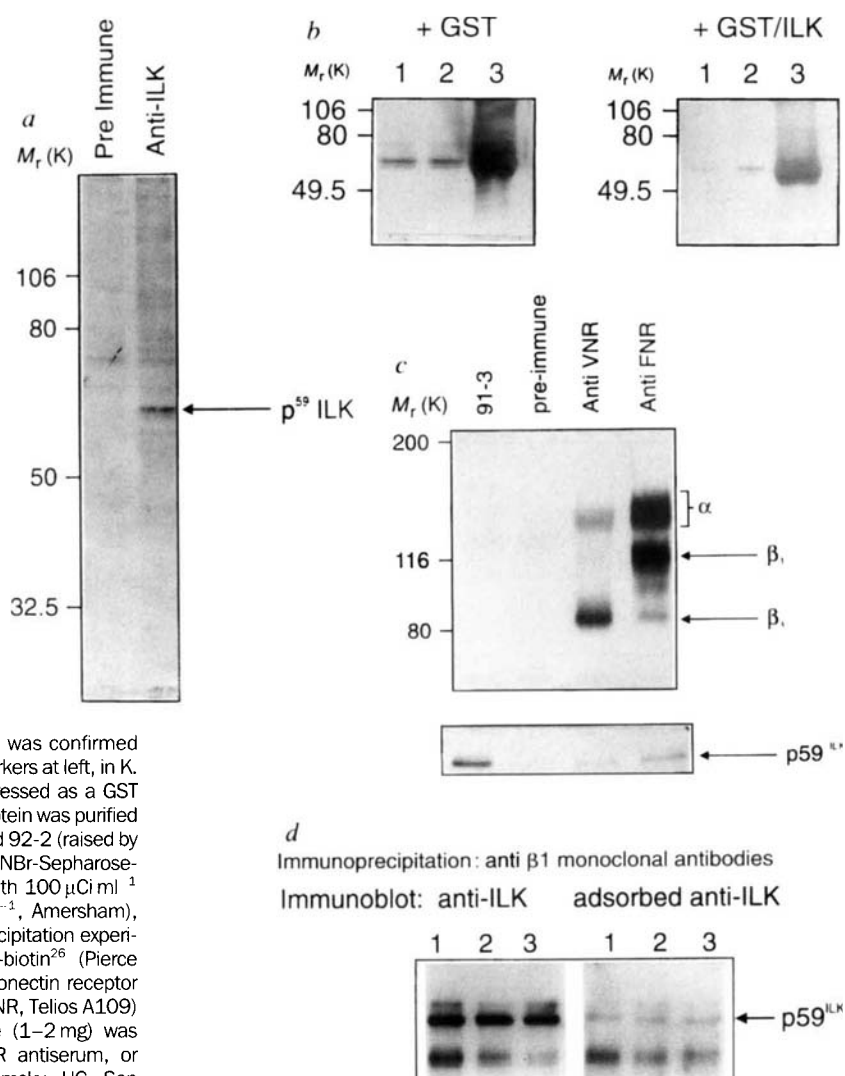
plasmids, pEG202 β_1 INT and pEG202 α_5 INT. pEG202 β_1 INT and pEG202 α_5 INT repressed β -gal expression from the pJK101 reporter by 50–60% and 70–75%, respectively, in host strain EGY48 (MAT α , his3, trp1, ura3-52, LEU2::pLEU2-LexAop6, constructed by Erica Golemis, Massachusetts General Hospital), confirming nuclear expression of the LexA fusions. Cotransformation of baits with the pSH18-34 reporter verified they were transcriptionally inert (not shown). A galactose-inducible HeLa cDNA interactor library was present on the TRP+ vector, pJG4-5 (constructed by J. Gyuris, MGH). For the β_1 interaction trap, EGY48 was transformed sequentially with pEG202 β_1 INT, pSH18-34 and pJG4-5, using the lithium acetate protocol²² (transformation efficiency = 5–6 $\times$ 10⁴ μ g⁻¹). Primary transformants (2 $\times$ 10⁶) were screened, of which 49 interacting clones were confirmed. The most frequent isolate (31/49) was a 700 bp insert, BIT-9. Retransformation of EGY48 with the BIT-9, pSH18-34, and pEG202 β_1 INT plasmids resulted in strong β -galactosidase expression, confirming the interaction. An identical screen, using pEG202 α_5 INT as bait, resulted in the isolation of 16 positives, none of which were represented in the set of 49 β_1 interactors. Trapped inserts were used to screen WM35 human melanoma λ gt10, and human placental λ gt11 cDNA libraries, using standard procedures²³. cDNA sequencing of multiple clones from each library was done using the dideoxy chain termination method (Sequenase 2.0, US Biochemical). For data analysis we used the Genetics Computer Group software package (version 7.0), and database searches were accomplished using the BLAST²⁸ server at the National Center for Biotechnology Information.

(ref. 12), on fibronectin reduced ILK phosphorylation of MBP in immune complex kinase assays, relative to cells plated on plastic, or kept in suspension (Fig. 4a). This fibronectin-dependent reduction of ILK activity was abrogated in IEC-18 cells expressing an activated *H-ras* allele¹², indicating that *ras* transformation disrupts extracellular matrix (ECM) regulation of ILK activity in these cells. An expression vector containing the full-length ILK cDNA, pCMV-ILK, was stably transfected into IEC-18 cells. Twelve stable clones each, of pCMV-ILK and vector control transfectants, were selected and characterized for p58^{ILK} expression levels. Two representative overexpressing subclones, ILK13-Ala3 and -A4a are illustrated (Fig. 4b). Overexpression of p59^{ILK} disrupted the epithelial morphology of IEC-18 cells. ILK13 clones were more refractile, and grew on laminin, fibronectin and vitronectin with a stellate morphology, in marked contrast to the typical 'cobble-stone' morphology of the parental and ILK14 cells (Fig. 4c). We plated the ILK13-Ala3 and -A4a subclones, the control transfectants, ILK14-A2C3 and -A2C6, and IEC-18 cells, on varying concentrations of the integrin substrates, laminin, fibronectin and vitronectin. Adhesion of the ILK14 and IEC-18 cells was equivalent, whereas that of the overexpressing subclones

was significantly reduced, on all these substrates (Fig. 4d). Immunoprecipitation analysis indicated that cell-surface integrin expression was unaffected (not shown). The effect of p59^{ILK} overexpression on anchorage-independent growth was examined by assaying the colony-forming ability of ILK transfectants in soft agarose. In marked contrast to IEC-18 and transfectant controls, four independent p59^{ILK} overexpressing subclones, ILK13-A4a, A1a3, A4d3 and A4C12, formed colonies in these assays (Fig. 4e). The proliferative rates of all of these clones on tissue culture plastic were equivalent to control rates. These results demonstrate that p59^{ILK} overexpression in the IEC epithelial cells provides a growth advantage, in the absence of proliferative signals normally provided by adhesion.

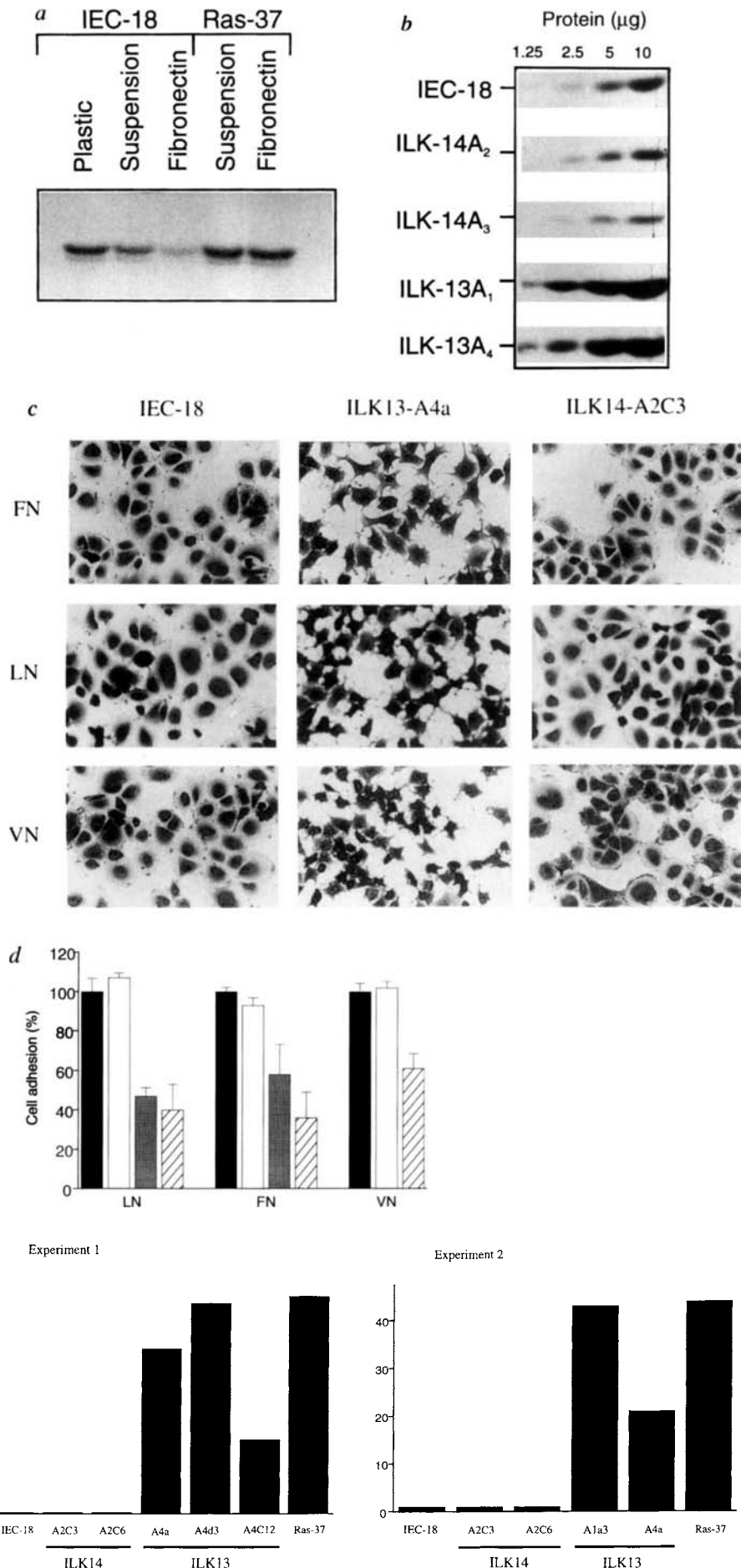
The transduction of extracellular matrix signals through integrins influences intracellular ('outside-in') and extracellular ('inside-out') functions, both of which appear to require interaction of integrin cytoplasmic domains with cellular proteins^{11,13}. The association of ILK with β -integrin subunits, and specific regulation of its kinase activity by adhesion to fibronectin, suggests that p59^{ILK} is a mediator of integrin signalling. Thus the ankyrin repeat motif probably represents a protein interaction module

FIG. 3 Antibodies to GST-ILK¹³² recognize p59^{ILK} in integrin coimmunoprecipitations. **a**, Unfractionated polyclonal anti-ILK sera 91-3 (shown) and 92-2 specifically recognize a ³⁵S-methionine, metabolically labelled cellular protein, of apparent *M_r* of 59K. A fluorograph is shown (En³Hance, NEN). **b**, Affinity-purified 92-2 antibody was adsorbed with 165 μ g of agarose-coupled GST-ILK¹³², or agarose-GST, and these preparations were used in parallel western blots containing 10 μ g per lane of whole-cell lysates of PC3 cells, Jurkat T-lymphoblasts, or the 60K GST-ILK¹³². **c**, Polyclonal anti-integrin antibodies, specific for the fibronectin and vitronectin receptors, were used to precipitate surface-biotinylated integrins from PC3 cells, and immune complexes were then analysed for the presence of p59^{ILK}, by western blotting with affinity-purified, biotin-labelled 92-2 antibody. This result is representative of six independent experiments. **d**, Anti- β_1 monoclonal antibodies were used in coprecipitation analyses of NP-40 lysates of PC3: lane 1, A₁B₂; lane 2, anti-CD29; lane 3, 3S3. Western blotting of anti- β_1 immune complexes with affinity-purified, biotinylated 92-2 antibody (left). This blot was stripped and reprobed with the same concentration of biotinylated 92-2, adsorbed against an excess of GST-ILK¹³² beads (right). We observe coprecipitation of p59^{ILK} using a panel of 11 anti- β_1 monoclonals, but not with an anti-CD44 monoclonal antibody (not shown). The migration of p59^{ILK} was confirmed in parallel lanes containing PC3 whole-cell NP-40 lysates. Markers at left, in K. **METHODS**. Amino-acid residues 132–451 of ILK were expressed as a GST fusion protein in *Escherichia coli*. Recombinant GST-ILK¹³² protein was purified and used to inject two rabbits. The resulting antisera, 91-3 and 92-2 (raised by Research Genetics), were affinity-purified over a column of CNBr-Sepharose-coupled GST-ILK¹³². PC3 cells were metabolically labelled with 100 μ Ci ml⁻¹ [³⁵S]methionine/[³⁵S]cysteine ([³⁵S] ProMix, 1,000 Ci mmol⁻¹, Amersham), for 18 h in cysteine/methionine-free MEM. For coimmunoprecipitation experiments PC3 cells were surface-labelled with sulfo-NHS-biotin²⁶ (Pierce Chemicals), before lysis in NP-40 buffer. Polyclonal anti-fibronectin receptor (anti-FNR, Telios A108), and anti-vitronectin receptor (anti-VNR, Telios A109) antibodies were purchased from Gibco/BRL. NP-40 lysate (1–2 mg) was incubated at 4 °C, with 2–3 μ l ml⁻¹ anti-FNR or anti-VNR antiserum, or 2 μ g ml⁻¹ the anti- β_1 monoclonal antibodies A₁B₂ (C. Damsky, UC, San Francisco), anti-CD29 (Upstate Biotechnology), and 3S3 (J. Wilkins, Univ. Manitoba). Lysates were precleared and immune complexes collected with Protein A-Sepharose. For western blotting, RIPA²⁷ lysates or immune complexes were subjected to 7.5% or 10% SDS-PAGE, and proteins then electrophoretically transferred to PVDF membranes (Immobilion-P, Millipore). Membranes were blocked in 5% non-fat milk/Tris-buffered saline Tween-20, and incubated with 0.5 μ g ml⁻¹ affinity-purified antibodies. Horseradish peroxidase-coupled goat anti-rabbit IgG was used in secondary incubations,



followed by detection of reactive bands by enhanced chemiluminescence (ECL, Amersham). For blotting without use of secondary antibody, affinity-purified 92-2 antibody was labelled with Biotin Hydrazide (Immunopure, Pierce Chemicals), according to the manufacturer's protocol, with visualization by peroxidase-conjugated streptavidin (Jackson ImmunoResearch Laboratories) and ECL. For reprobing, membranes were stripped according to manufacturer's instructions.

FIG. 4 Modulation of ILK kinase activity by ECM components. **a**, ILK phosphorylation of MBP was assayed in ILK immune complexes from lysates of IEC-18 intestinal epithelial cells which were collected from tissue culture plastic and either kept in suspension, or replated on fibronectin, for 1 h. An H-ras-transformed variant of IEC-18¹², Ras37 (transfected with Ras^{val12} in pRC/CMV vector), was assayed in parallel. The band shown is MBP. **b**, Expression levels of p59^{ILK} in two representative clones of IEC-18 cells, transfected with an ILK-expression construct (ILK13), two vector control clones (ILK14), and the parental IEC-18 cells. The indicated amounts (μg per lane) of whole-cell RIPA lysates were run out on 10% SDS-PAGE gels, and p59^{ILK} expression analysed by western blotting with affinity-purified 92-2 antibody. **c**, Representative p59^{ILK} overexpressing clone ILK13-A4a, vector control clone ILK14-A2C3, and parental IEC-18 cells were plated on the ECM substrates (aminin (LN) fibronectin (FN) vitronectin (VN) and for 1 h, then fixed, stained with toluidine blue and photographed ($\times 40$ magnification). **d**, Adhesion of the ILK overexpressing clones to LN, FN and VN was quantified, Key, IEC-18 (black), ILK14-A2C6 (white), ILK13-A4a3 (grey), ILK13-A4a (hatched). Results are presented for $10 \mu\text{g ml}^{-1}$ substrate, and are expressed as % adhesion ($\pm$ s.d.) relative to IEC-18, for each substrate. The serial concentrations of ECM showed similar reductions in adhesion of the ILK13 subclones, and ILK14-A2C3 adhesion was identical to that of ILK14-A2C6, on all three substrates. Immunoprecipitation of surface-biotinylated IEC-18, ILK13 and ILK14 subclones, with the anti-FNR and anti-VNR sera, confirmed there as no change in expression of $\alpha_5/\alpha_3\beta_1$ and $\alpha_v\beta_3/\beta_5$ integrin subunits in the p59^{ILK} overexpressors (data not shown). Data are representative of two independent experiments. **e**, Four ILK13, p59^{ILK} overexpressing clones were plated in soft agarose, and assayed for colony growth after three (experiment 1) and two (experiment 2) weeks. Parent and vector control transfectants were also assayed, and the rats^{val12} transformed clone, Ras-37, was used as a positive control. Bars represent the mean of duplicate determinations. Maximum colonies in IEC-18 and ILK14 cells was 1 per field. **METHODS.** The rat intestinal epithelial cell line IEC-18, and a variant of this line transfected with an activated H-ras^{val12} allele, expressed from pRC/CMV, were grown on tissue culture plastic in 5% serum-containing medium, washed three times in minimum essential medium (MEM), and collected with 5 mM EDTA. These were resuspended in 2.5 mg ml^{-1} BSA in MEM, and either kept in suspension, or plated on $10 \mu\text{g ml}^{-1}$ fibronectin-coated plates, for 1 h at 37°C . NP-40 lysates ($300 \mu\text{g}$) of these cells were immunoprecipitated with affinity-purified 91-3, and immune complex kinase assays (MBP substrate) performed, as described above. IEC-18 were transfected with the expression vector pRC/CMV, containing Plac5 in the forward orientation relative to the CMV promoter. Stable clones were selected in G418, and subcloned through two rounds of limiting dilution. In all, 12 each of ILK and vector control transfectant subclones were isolated. Protein concentrations were determined using the Bradford reagent (Bio-Rad). Two p59^{ILK} overexpressors, ILK13-A4a3 and ILK13-A4a, and two vector transfectant controls, ILK14-A2C3 and -A2C6, were analysed for effects of ILK overexpression on cell adhesion to ECM substrates. Adhesion was quantified according to published methods²⁸. For colony formation assays 3×10^5 cells were plated in 35-mm wells, in 0.3% agarose, as described previously²⁹. Ras-37 were plated at 2×10^3 per well. Colonies were counted and scored per field ($d = 1 \text{ cm}$) in duplicate wells, and defined as a minimum aggregate of 50 cells.



specifying interactions of ILK with downstream, cytoplasmic or cytoskeletal proteins. Reduced ECM adhesion by the p59^{ILK} overexpressing cells is consistent with our observation of adhesion-dependent inhibition of ILK activity, and suggests that p59^{ILK} plays a role in inside-out integrin signalling. Furthermore the p59^{ILK}-induced, anchorage-independent growth of epithelial cells indicates a role for ILK in mediating intracellular signal transduction by integrins¹⁴⁻¹⁶. □

Received 15 August; accepted 30 October 1995.

1. Damsky, C. H. & Werb, Z. *Curr. Opin. Cell Biol.* **4**, 772-781 (1992).
2. Hynes, R. O. *Cell* **69**, 11-25 (1992).
3. Clark, E. A. & Brugge, J. S. *Science* **268**, 233-239 (1995).
4. Fields, S. & Song, O. *Nature* **340**, 245-246 (1989).
5. Lux, S. E., John, K. M. & Bennett, V. *Nature* **344**, 36-42 (1990).
6. Inoue, J.-I. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 4333-4337 (1992).
7. Lukas, J. et al. *Nature* **375**, 503-506 (1993).
8. Schaller, M. D. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 5192-5196 (1992).
9. Hanks, S. K., Calalb, M. B., Harper, M. C. & Patel, S. K. *Proc. natn. Acad. Sci. U.S.A.* **89**, 8481-8491 (1992).
10. Dedhar, S., Saulnier, R., Nagle, R. & Overall, C. M. *Clin. exp. Metastasis* **11**, 391-400 (1993).
11. Chen, Y.-P. et al. *J. biol. Chem.* **269**, 18307-18310 (1994).
12. Filmus, J. et al. *Oncogene* **9**, 3627-3633 (1994).
13. O'Toole, T. E. et al. *J. Cell Biol.* **124**, 1047-1059 (1994).
14. Kapron-Bras, C., Fitz-Gibbon, L., Jeevaratnam, P., Wilkins, J. & Dedhar, S. *J. biol. Chem.* **268**, 20701-20704 (1993).

15. Chen, Q., Kinch, M. S., Lin, T. H., Burrige, K. & Juliano, R. L. *J. biol. Chem.* **269**, 26602-26605 (1994).
16. Schlaepfer, D. D., Hanks, S. K., Hunter, T. & van der Geer, P. *Nature* **372**, 786-791 (1994).
17. Kozak, M. *Cell* **44**, 283-292 (1986).
18. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. *J. molec. Biol.* **215**, 403-410 (1990).
19. Hanks, S. K., Quinn, A. M. & Hunter, T. *Science* **241**, 42-52 (1988).
20. Zeros, A. S., Gyuris, J. & Brent, R. *Cell* **72**, 223-232 (1993).
21. Argraves, W. S. et al. *J. Cell Biol.* **105**, 1183-1190 (1987).
22. Gietz, D., St. Jean, A., Woods, R. A. & Schiestl, R. H. *Nucleic Acids Res.* **20**, 1425 (1992).
23. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1989).
24. Otey, C. A., Pavalko, F. M. & Burridge, K. *J. Cell Biol.* **111**, 721-729 (1990).
25. Cooper, J. A., Sefton, B. M. & Hunter, T. *Meth. Enzym.* **99**, 387-402 (1983).
26. Stephens, L. C., Sonne, J. E., Fitzgerald, M. L. & Damsky, C. H. *J. Cell Biol.* **123**, 1607-1620 (1993).
27. Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1988).
28. Leung-Hagstegen, C. Y., Milankov, K., Michalak, M., Wilkins, J. & Dedhar, S. *J. Cell Sci.* **107**, 589-600 (1994).
29. Buick, R. N., Filmus, J. & Quaroni, A. *Exp. Cell Res.* **170**, 300-309 (1987).

ACKNOWLEDGEMENTS. We are grateful to R. Brent and R. Finley (Massachusetts General Hospital) for providing the yeast strains and two-hybrid plasmids, and to A. Murthy (MGH) for discussions regarding yeast transformations. Thanks to C. Otey (Univ. of Virginia) for providing the synthetic integrin cytoplasmic domain peptides, Y. Ben-David and S.-J. Lu (Sunnybrook) for the WM35 melanoma cDNA library, and M. Park (McGill University) for the placental cDNA library. Monoclonal antibodies to β_1 integrin were provided by C. Damsky (Univ. of California at San Francisco) and J. Wilkins (Univ. of Manitoba). We appreciate critical reading of the manuscript by J. Woodgett (Univ. of Toronto). G.H. was the recipient of postdoctoral fellowships from the Medical Research Council of Canada and the Sunnybrook Trust for Medical Research. S.D. is a Terry Fox Scientist of the National Cancer Institute of Canada, and is supported by grants from the NCIC and MRC(C).

Conserved residues and the mechanism of protein folding

E. Shakhnovich, V. Abkevich & O. Ptitsyn*

Harvard University, Department of Chemistry, 12 Oxford Street, Cambridge, Massachusetts 02138, USA

*Institute of Protein Research, Russian Academy of Sciences, Puschino, Moscow Region 142292, Russia, and NIH, National Cancer Institute, Molecular Structure Section, LMMB, Bethesda, Maryland 20892-3677, USA

EXPERIMENTAL¹⁻⁶ and simulation⁷ studies show that small monomeric proteins fold in one kinetic step, which entails overcoming the free-energy barrier between the unfolded and the native protein through a transition state^{8,9}. Two models of transition state formation have been proposed: a 'nonspecific' one in which it depends on the formation of a sufficient number of native-like contacts regardless of what amino acids are involved¹⁰⁻¹², and a 'specific' one, in which it depends on formation of a specific subset of the native structure (a folding nucleus)^{8,13,14}. The latter requires that some amino acids form most of their contacts in the transition state, whereas others only do so on reaching the native conformation. If so, mutations affecting the stability of the transition state nucleus should have a greater effect on the folding kinetics than mutations elsewhere, and the residues involved should be evolutionarily conserved. Lattice-model simulations and experiments^{8,13-16} suggest that such mutations exist. Here we present a method for determining the folding nucleus of a protein with known structure with two-state folding kinetics. This method is based on the alignment of many sequences designed to fold into the native conformation of a protein to identify the positions where amino acids are most conserved in designed sequences. The method is applied to chymotrypsin inhibitor 2 (CI2), a protein whose transition state has been previously studied by protein engineering¹⁴⁻¹⁶. The involvement of residues in folding nucleus of CI2 is clearly correlated with their conservation in design, and the residues forming the nucleus are highly conserved in 23 natural sequences homologous to CI2.

We first studied a simple lattice model of the protein¹⁷. (1) We chose an arbitrary conformation of lattice protein chain to serve as the native structure and then (2) selected sequences that deliver low energy in this 'native' structure compared to unfolded and

misfolded conformations. (3) The designed sequences (from 2) were folded using Monte Carlo simulations. Because folding simulations and sequence design are carried out using the same set of potentials, a self-consistent study of the model can be carried out with any choice of potential. For some (but not all) predictions on real proteins it is possible that the particular choice of potentials is not very important¹⁷ (see below).

Following this method we chose the target conformation (Fig. 1), and designed sequences to fold to it. Further, the analysis based on Monte Carlo folding simulations⁸ revealed the folding nucleus (Fig. 1a) for both sequences shown in Fig. 1b,c.

The stochastic design algorithm generated 10⁶ sequences, each having low energy in the conformation shown in Fig. 1a. Alignment of these sequences revealed a remarkable feature: correlation between residue conservation and its participation in the folding nucleus (Fig. 2). Indeed, all four most conserved residues (5,16,20,35) belong to the nucleus (see Fig. 1a). This could have been due to the degree they are buried in the native structure (see Fig. 1), but this is not the case because the conservation of all buried residues is smaller than that of nucleus residues (see Fig. 2b).

The key feature of the nucleation mechanism is the existence of 'kinetically important' positions. To test this we designed a set of sequences where positions (5,16,20,35) inside the nucleus are constrained to 'alanines', which destabilize the nucleus (in an MJ parameter set). We studied the folding kinetics for these sequences (Fig. 3). The existence of some 'alanine' residues in the nucleus caused a pronounced slowing down of folding. This is an exclusively kinetic effect: several 'wild-type' sequences (designed without forcing 'alanines' into the nucleus) have native-state energy comparable to that of sequences designed with alanines in the nucleus. This example implies that the present method of identifying a folding nucleus would give wrong predictions for sequences with alanines designed in the nucleus, because these sequences do not appear to fold through a specific nucleus mechanism. Therefore the present method only applies to fast-folding sequences.

We applied this method to predict the folding nucleus of CI2, originally studied in refs 15, 16. To this end we used the real off-lattice structure of CI2 as the target and designed sequences that have low energy in this conformation. There is a clear qualitative correlation between site conservation and ϕ -values for folding. Our calculations (see Fig. 4) predict that most conserved residues, particularly A35, I39, L68, I70 and I76, are likely to belong to the folding nucleus. After this work was completed, we learned from A. Fersht (personal communication) that A35 (not studied in refs

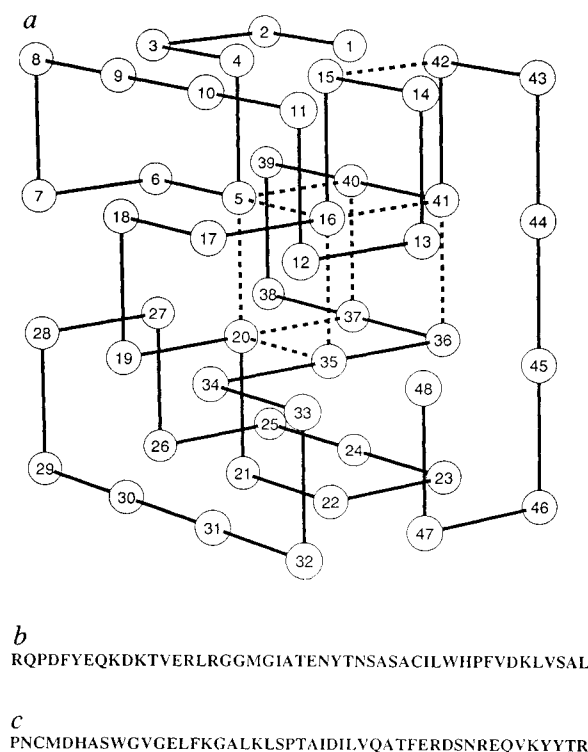


FIG. 1 a, The conformation of the 48-mer chosen as the native state in our design/folding procedure. Each amino-acid residue is represented as a bead occupying a lattice site. Although the model does not treat side chains explicitly, the amino acids are chemically different; their differences are manifested in pairwise interaction energies of different magnitude and sign, depending on the identities of interacting amino acids. A conformation is described by the set of coordinates of all monomers $\{r_i\}$. The energy of a conformation is:

$$E = \sum_{i < j} B(\eta_i, \eta_j) \Delta(r_i - r_j) \quad (1)$$

where $\Delta(r_i - r_j) = 1$ if monomers i and j are lattice neighbours not connected by a covalent bond, and 0 otherwise. $B(\eta_i, \eta_j)$ is the magnitude of interaction between amino acids of type η_i and η_j . There are 20 types of amino acid in the model. Two parameter sets, B , were used: one proposed by Miyazawa and Jernigan¹⁹ (MJ), and the other proposed by Kolinski, Godzik and Skolnick²⁰ (KGS). The parameters given in Table 6 of ref. 19 were used as the MJ set. They represent the 'excess' pairwise interaction between two amino acids as compared with their averaged interaction with their environment. The values of B in equation (1) were shifted and normalized to achieve a zero average over all possible contacts and standard variance of unity. This amounts to multiplying all parameters in refs 19, 20 by a constant factor and adding a constant to each interaction parameter B . This procedure effectively sets the temperature scale for simulations (more details in ref. 8). For each set of parameters we designed sequences to fold to the conformation shown here. The design procedure has been described in detail elsewhere^{17,21,22}. It is a stochastic (Monte Carlo) optimization routine in sequence space which keeps amino-acid composition unchanged. It minimizes energy of the native conformation. The condition of constant amino-acid composition makes it equivalent to optimizing the relative energy of the native state, or Z-score²³. As is characteristic of Monte Carlo searches, unfavourable mutations can also be accepted, with a small probability, given by a Metropolis criterion²⁴ with selective temperature T_{sel} . We chose $T_{\text{sel}} = 0.15$ (in our temperature scale), which is sufficiently low to generate stable and fast-folding sequences. b, c, Two sequences designed to fold and be stable in the conformation shown in a: with KGS parameters (b), and with MJ parameters (c). The design tends to place the most strongly interacting amino acids in the interior where they can form most contacts. The strongest 'excess' interactions in MJ parameters from ref. 19 are between 'charged' groups (D and K) therefore they are buried in sequence (c). Alternatively, the strongest interactions in the KGS set are between hydrophobic groups, hence designs with these parameters yield more realistic sequences with hydrophobic groups buried, as in sequence b. Monte Carlo folding simulations were carried out for both sequences. Both of them folded to and were stable in the conformation shown in a with their respective forcefields. The nucleus (determined from the folding simulations, using the method described in detail in ref. 8) was identical for both sequences; it is shown by broken lines in a.

15, 16) is the residue most involved in the nucleus with $\phi \approx 1.0$ (see also ref. 14); I76 looks like an exception with $\phi \approx 0.13$. However, it is not: despite the low ϕ -value, recent measurements on a number of residues contacting I76 strongly suggest that I76 is also a key residue of the folding nucleus¹⁴.

We also found a striking relationship between the residue conservatism in our design and evolutionary conservatism in the alignment of natural sequences homologous to CI2 (Fig. 4b)¹⁸. The key nucleus residue A35 is 100% conserved. Among the most conserved in Fig. 4b are nucleus residues I76, I39. Nucleus L68 is moderately conserved in the alignment, though, because of more frequent substitutions of L to V.

This finding also suggests that our design procedure might reflect on some features of the evolutionary process of protein morphogenesis, namely these aspects related to folding. It is possible that some of the surface residues are evolutionarily conserved for functional reasons not taken into account in the design procedure. To this end, the comparison of the results of the design (Fig. 4a) with the alignment (Fig. 4b) might be helpful in identifying which residues are conserved for 'folding' reasons and which ones are functionally conserved.

The possible physical rationale for the suggested method is that the sequence design identifies a contiguous cluster of core residues which are close enough to each other in space to form a contiguous network of interactions. The energy-based design

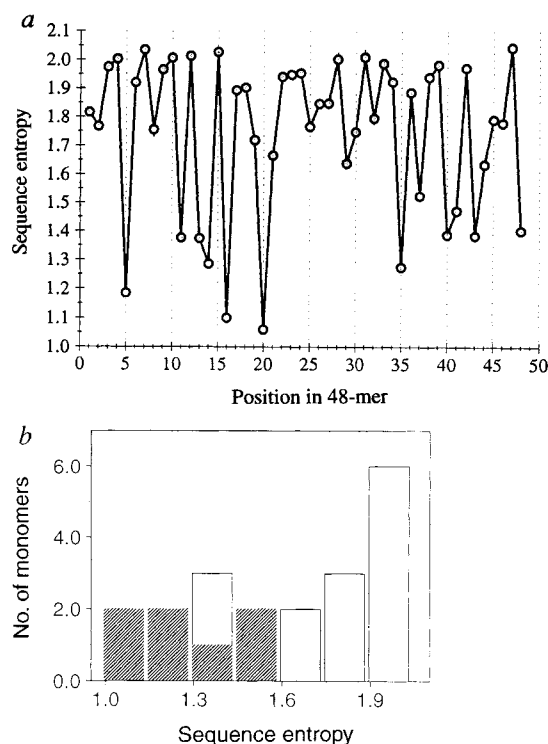


FIG. 2 a, The sequence design entropy is defined as:

$$S(i) = - \sum_{j=1}^m p_j(i) \ln p_j(i) \quad (2)$$

where $p_j(i)$ is frequency of occurrence of amino acid of type j at site i ; $m = 20$ is the total number of amino-acid types. To estimate the frequencies $p_j(i)$, we performed long runs of the Monte Carlo sequence design algorithm using a low selective temperature ($T_{\text{sel}} = 0.15$) to obtain $\sim 10^6$ sequences per parameter set. We estimated the desired frequencies as the fraction of this population of designed sequences bearing an amino acid of type j at position i . Shown is a plot of $S(i)$ as a function of position i for the KGS parameter set; the corresponding plot for the MJ parameter set is similar (not shown). b, Histogram for the distribution of design entropy for buried residues (having three or four non-covalent contacts in the native structure, Fig. 1a) from the plot in a. White bars, all buried residues; grey bars, those of buried residues that have two or more nucleus contacts.

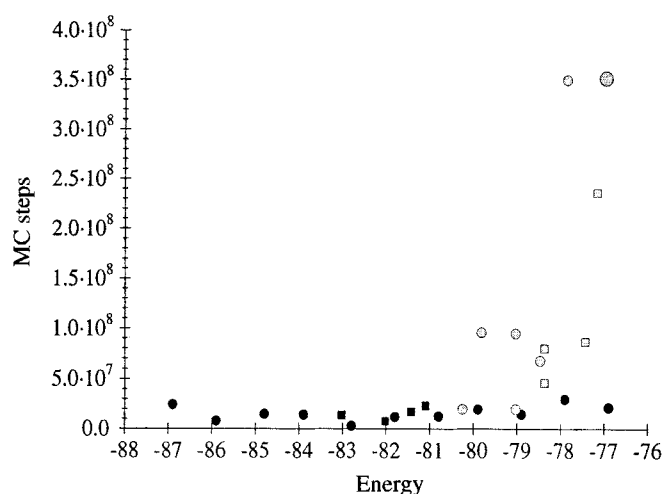


FIG. 3 The mean first-passage folding time for different sequences designed (with MJ parameters) to fold into the native structure shown in Fig. 1a. Monte Carlo folding simulations were done at a low temperature ($T = 0.8$) so that the folding time of the 'wild-type' sequences (black circles) would not depend dramatically on native-state energy²⁵. This is indeed the case in the present simulations: 11 'wild-type' sequences having different energies in the native state (black circles) have similar average folding times. Low temperature for folding simulations was chosen to distinguish the special role of nucleus residues in folding from the more obvious²⁶⁻²⁹ relation between folding rate and the stability of the native state. In the parameters from ref. 19, alanine interacts unfavourably with most other amino acids so that its placement in the nucleus destabilizes the latter. Several sequences were designed to have an alanine residue at the predetermined nucleus positions. (Alanine residues were placed at these positions and mutations were forbidden there, otherwise the design algorithm proceeded as usual.) Black squares correspond to sequences with one alanine residue, placed in either positions 5,16,20,35; grey circles correspond to sequences with 2 fixed alanine residues, in positions (5,16), (5,20), (16,20), (5,35), (16,35) or (20,35); grey squares correspond to sequences having 3 fixed alanines placed in all triplets out of positions (5,16,20,35), and the large grey circle corresponds to the sequence designed with fixed alanine residue at all four positions (5,16,20,35).

procedure conservatively places into these positions residues that strongly attract each other. This creates a relatively low free-energy set of partly folded conformations in which mutually stabilizing strong nucleus contacts are formed while other parts of the chain are disordered. In the model representing folding as kinetically two-state, such a set of conformations serves as a saddle-point in the free-energy landscape, which is the transition state. This also suggests that our current method may be applicable only to proteins with two-state folding kinetics.

Note added in proof: The notation for C12 residues in ref. 14 is shifted by 19 units from the notation used in ref. 15 and this work; for example, A35 here would be A16 in ref. 14. □

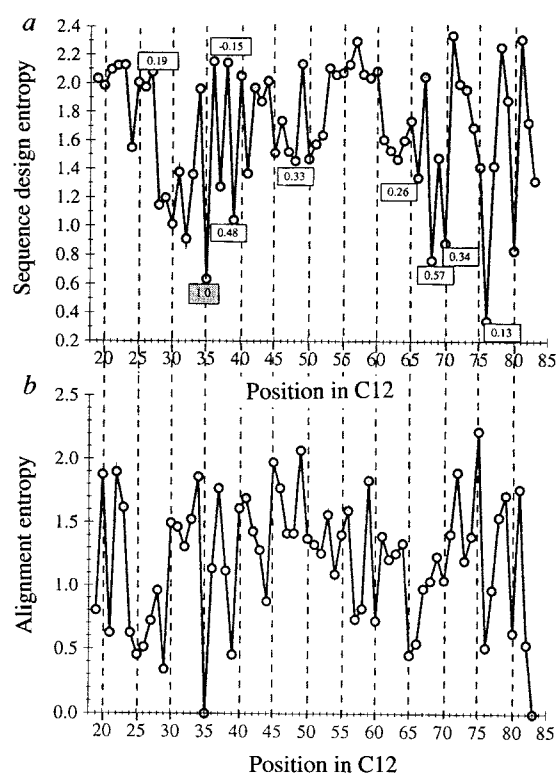


FIG. 4 The design entropy of C12. The PDB-structure of C12 (2ci2) was taken as the target conformation. Then, using an MC sequence design algorithm, we generated many sequences (with the same amino-acid composition as the native C12) exhibiting a low energy in this target conformation. The energy function for design was calculated for C_{β} contact using equation (1). Two residues were considered to be in contact if the distance between their C_{β} atoms (C_{α} for G) was ≤ 7.5 Å and if they were more than two units apart from each other along the sequence. KGS and MJ parameters were used for $B(\eta_i, \eta_j)$ (see equation (1)). Amino-acid frequencies were taken from the designed sequences (as explained in Fig. 2) and substituted into equation (2). We compared these results with the numbers of native-like contacts of a given residue in the transition state relative to that in the native state (ϕ -values^{30,31}). Labels represent the ϕ -values for all residues studied in refs 15, 16. The shaded label is the recent result of ref. 14 which only came to our attention after this study was completed. The plot shown here was calculated using the KGS parameters. The MJ parameters yielded very similar results indicating the same positions as conservative ones. Design with KGS parameters placed predominantly L and I in the conserved positions, which coincided in most cases (except position 35) with what is seen in real sequences. MJ parameters placed charged groups in these positions, for the reasons explained in the legend to Fig. 1. b, The alignment entropy (amino-acid variability) calculated over 23 sequences homologous to C12 (ref. 18). Frequencies of amino-acid occurrences were taken from the alignment, and sequence alignment entropy was evaluated according to equation (2). (These data were already calculated in the file we used, 2ci2.hssp (ref. 18).)

Received 25 July; accepted 2 November 1995.

- Jackson, S. E. & Fersht, A. R. *Biochemistry* **30**, 10428–10435 (1991).
- Alexander, P., Orban, J. & Bryan, P. *Biochemistry* **31**, 7243–7248 (1992).
- Schindler, T., Herfner, M., Marahiel, M. & Schmid, F. X. *Nature struct. Biol.* **2**, 663–673 (1995).
- Sosnick, T. R., Mayne, L., Hiller, R. & Englander, S. W. *Nature struct. Biology* **1**, 149–156 (1994).
- Viguera, A. R., Martinez, J. C., Filiminov, V. V., Mateo, P. L. & Serrano, L. *Biochemistry* **33**, 2142–2150 (1994).
- Kragelund, B. B., Robinson, C. V., Knudsen, J. & Dobson, C. M. *Biochemistry* **34**, 7117–7124 (1995).
- Gutin, A. M., Abkevich, V. I. & Shakhnovich, E. I. *Biochemistry* **34**, 3066–3076 (1995).
- Abkevich, V. I., Gutin, A. M. & Shakhnovich, E. I. *Biochemistry* **33**, 10026–10036 (1994).
- Guo, Z. & Thirumalai, D. *Biopolymers* **35**, 137–139 (1995).
- Shakhnovich, E. I., Farztdinov, G., Gutin, A. M. & Karplus, M. *Phys. Rev. Lett.* **67**, 1665–1668 (1991).
- Sali, A., Shakhnovich, E. I. & Karplus, M. *Nature* **369**, 248–251 (1994).
- Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. *Science* **267**, 1619–1620 (1995).
- Fersht, A. *Proc. natn. Acad. Sci. U.S.A.* **92**, 10869–10873 (1995).
- Itzhaki, L., Oltzen, D. & Fersht, A. *J. molec. Biol.* **254**, 260–288 (1995).
- Jackson, S. E., elMasry, N. & Fersht, A. *Biochemistry* **32**, 11270–11278 (1993).
- Oltzen, D. E., Itzhaki, L., elMasry, N., Jackson, S. E. & Fersht, A. *Proc. natn. Acad. Sci. U.S.A.* **91**, 10422–10425 (1994).
- Shakhnovich, E. I. *Phys. Rev. Lett.* **72**, 3907–3910 (1994).

- Sander, C. & Schneider, R. *Proteins* **9**, 56–68 (1991).
- Miyazawa, S. & Jernigan, R. *Macromolecules* **18**, 534–552 (1985).
- Kolinski, A., Godzik, A. & Skolnick, J. J. *chem. Phys.* **98**, 7420–7433 (1993).
- Shakhnovich, E. I. & Gutin, A. M. *Proc. natn. Acad. Sci. U.S.A.* **90**, 7195–7199 (1993).
- Shakhnovich, E. I. & Gutin, A. M. *Proc. Engng* **6**, 793–800 (1993).
- Bowie, J. U., Luthy, R. & Eisenberg, D. *Science* **253**, 164–169 (1991).
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N. & Teller, E. *J. chem. Phys.* **21**, 1087–1092 (1953).
- Abkevich, V. I., Gutin, A. M. & Shakhnovich, E. I. *J. chem. Phys.* **101**, 6052–6062 (1994).
- Goldstein, R., Luthy-Schultzen, Z. A. & Wolynes, P. G. *Proc. natn. Acad. Sci. U.S.A.* **89**, 4918–4922 (1992).
- Sali, A., Shakhnovich, E. I. & Karplus, M. *J. molec. Biol.* **235**, 1614–1636 (1994).
- Hao, M.-H. & Scheraga, H. J. *chem. Phys.* **102**, 1334–1339 (1995).
- Gutin, A. M., Abkevich, V. I. & Shakhnovich, E. I. *Proc. natn. Acad. Sci. U.S.A.* **92**, 1282–1286 (1995).
- Matouschek, A., Kellis, J., Serrano, L., Bycroft, M. & Fersht, A. *Nature* **346**, 440–445 (1990).
- Matouschek, A., Serrano, L. & Fersht, A. *J. molec. Biol.* **224**, 819–835 (1992).

ACKNOWLEDGEMENTS. We are grateful to A. Fersht and M. Karplus for discussions and to G. Berriz for comments and editorial help. E.S. is supported by the Packard foundation. O.P. acknowledges support from the International Human Frontier Science Program. The alignment file 2ci2.hssp was obtained from the EMBL server, Heidelberg, Germany.